

## University of Groningen

### The first power spectrum limit on the 21-cm signal of neutral hydrogen during the Cosmic Dawn at $z=20-25$ from LOFAR

Gehlot, B. K.; Mertens, F. G.; Koopmans, L. V. E.; Brentjens, M. A.; Zaroubi, S.; Ciardi, B.; Ghosh, A.; Hatef, M.; Iliev, I. T.; Jelic, V.

*Published in:*  
Monthly Notices of the Royal Astronomical Society

*DOI:*  
[10.1093/mnras/stz1937](https://doi.org/10.1093/mnras/stz1937)

**IMPORTANT NOTE:** You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2019

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Gehlot, B. K., Mertens, F. G., Koopmans, L. V. E., Brentjens, M. A., Zaroubi, S., Ciardi, B., Ghosh, A., Hatef, M., Iliev, I. T., Jelic, V., Kooistra, R., Krause, F., Mellema, G., Mevius, M., Mitra, M., Offringa, A. R., Pandey, V. N., Sardarabadi, A. M., Schaye, J., ... Yatawatta, S. (2019). The first power spectrum limit on the 21-cm signal of neutral hydrogen during the Cosmic Dawn at  $z=20-25$  from LOFAR. *Monthly Notices of the Royal Astronomical Society*, 488(3), 4271-4287. <https://doi.org/10.1093/mnras/stz1937>

#### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

#### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

# The first power spectrum limit on the 21-cm signal of neutral hydrogen during the Cosmic Dawn at $z = 20\text{--}25$ from LOFAR

B. K. Gehlot<sup>1</sup>,<sup>1</sup>★ F. G. Mertens<sup>1</sup>,<sup>1</sup> L. V. E. Koopmans<sup>1</sup>,<sup>1</sup>★ M. A. Brentjens<sup>2</sup>,  
S. Zaroubi<sup>1,3</sup>, B. Ciardi<sup>4</sup>, A. Ghosh<sup>5,6</sup>, M. Hatef<sup>1,2</sup>, I. T. Iliev<sup>7</sup>, V. Jelić<sup>8</sup>, R. Kooistra<sup>1</sup>,  
F. Krause<sup>1,9</sup>, G. Mellema<sup>10</sup>, M. Mevius<sup>2</sup>, M. Mitra<sup>1</sup>, A. R. Offringa<sup>2</sup>, V. N. Pandey<sup>1,2</sup>,  
A. M. Sardarabadi<sup>1</sup>, J. Schaye<sup>11</sup>, M. B. Silva<sup>1</sup>, H. K. Vedantham<sup>2</sup> and S. Yatawatta<sup>1,2</sup>

<sup>1</sup>Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700AV Groningen, the Netherlands

<sup>2</sup>ASTRON, Netherlands Institute for Radio Astronomy, Oude Hoogeveensedijk 4, NL-7991 PD, Dwingeloo, the Netherlands

<sup>3</sup>Department of Natural Sciences, The Open University of Israel, 1 University Road, PO Box 808, Ra'anana 4353701, Israel

<sup>4</sup>Max-Planck Institute for Astrophysics, Karl-Schwarzschild-Straße 1, D-85748 Garching, Germany

<sup>5</sup>Department of Physics, University of the Western Cape, Cape Town 7535, South Africa

<sup>6</sup>SKA South Africa, 3rd Floor, The Park, Park Road, Pinelands, 7405 South Africa

<sup>7</sup>Astronomy Centre, Department of Physics and Astronomy, Pevensey II Building, University of Sussex, Brighton BN1 9QH, UK

<sup>8</sup>Ruder Bošković Institute, Bijenička cesta 54, 10000 Zagreb, Croatia

<sup>9</sup>Department of Physics and Astronomy, University College London, Gower Place, London WC1E 6BT, UK

<sup>10</sup>Department of Astronomy and Oskar Klein Centre for Cosmoparticle Physics, AlbaNova, Stockholm University, SE-106 91 Stockholm, Sweden

<sup>11</sup>Leiden Observatory, Leiden University, PO Box 9513, NL-2300RA Leiden, the Netherlands

Accepted 2019 July 8. Received 2019 July 2; in original form 2018 September 18

## ABSTRACT

Observations of the redshifted 21-cm hyperfine line of neutral hydrogen from early phases of the Universe such as Cosmic Dawn and the Epoch of Reionization promise to open a new window onto the early formation of stars and galaxies. We present the first upper limits on the power spectrum of redshifted 21-cm brightness temperature fluctuations in the redshift range  $z = 19.8\text{--}25.2$  (54–68 MHz frequency range) using 14 h of data obtained with the LOFAR-Low Band Antenna (LBA) array. We also demonstrate the application of a multiple pointing calibration technique to calibrate the LOFAR-LBA dual-pointing observations centred on the North Celestial Pole and the radio galaxy 3C220.3. We observe an unexplained excess of  $\sim 30\text{--}50$  per cent in Stokes  $I$  / noise compared to Stokes  $V$  for the two observed fields, which decorrelates on  $\gtrsim 12$  s and might have a physical origin. We show that enforcing smoothness of gain errors along frequency direction during calibration reduces the additional variance in Stokes  $I$  compared Stokes  $V$  introduced by the calibration on sub-band level. After subtraction of smooth foregrounds, we achieve a  $2\sigma$  upper limit on the 21-cm power spectrum of  $\Delta_{21}^2 < (14561 \text{ mK})^2$  at  $k \sim 0.038 h \text{ cMpc}^{-1}$  and  $\Delta_{21}^2 < (14886 \text{ mK})^2$  at  $k \sim 0.038 h \text{ cMpc}^{-1}$  for the 3C220 and NCP fields respectively and both upper limits are consistent with each other. The upper limits for the two fields are still dominated by systematics on most  $k$  modes.

**Key words:** methods: data analysis – methods: statistical – techniques: interferometric – dark ages, reionization, first stars – diffuse radiation – radio lines: general.

## 1 INTRODUCTION

After the Epoch of Recombination around redshift  $z \sim 1100$ , the Universe entered the ‘Dark Ages’ era during which it was completely neutral and devoid of any radiation sources. During

this period, small perturbations in matter density grew under gravitational instability, and matter started to accumulate in localized overdensity peaks. The formation of the first luminous objects (stars and galaxies) in these overdense regions marked the beginning of the so-called Cosmic Dawn (CD) era spanning the redshift range  $30 > z > 12$  (Pritchard & Furlanetto 2007). X-ray and ultraviolet radiation from the first stars and galaxies began to heat and ionize the neutral hydrogen (HI hereafter) in the surrounding intergalactic medium (IGM), starting off the Epoch of Reionization (EoR,  $12 >$

\* E-mail: [kbharatgehlott@gmail.com](mailto:kbharatgehlott@gmail.com) (BGK); [koopmans@astro.rug.nl](mailto:koopmans@astro.rug.nl) (LVEK)

$z > 6$ ) during which HI in the IGM transitioned from being fully neutral to ionized (Madau, Meiksin & Rees 1997).

The redshifted 21-cm signal corresponding to the hyperfine transition of HI has been identified as an excellent probe of the HI distribution in the IGM during the CD and the EoR (Madau et al. 1997; Shaver et al. 1999; Furlanetto, Oh & Briggs 2006; Pritchard & Loeb 2012; Zaroubi 2013). A number of ongoing and upcoming experiments, such as the LOw Frequency ARray<sup>1</sup> (LOFAR; van Haarlem et al. 2013), the Giant Meterwave Radio Telescope<sup>2</sup> (GMRT; Paciga et al. 2011), the Murchison Widefield Array<sup>3</sup> (MWA; Tingay et al. 2013; Bowman et al. 2013), the Precision Array for Probing the Epoch of Reionization<sup>4</sup> (PAPER; Parsons et al. 2010), the Hydrogen Epoch of Reionization Array<sup>5</sup> (HERA; DeBoer et al. 2017), NenuFAR<sup>6</sup> (New extension in Nançay Upgrading loFAR; Zarka et al. 2012), and the Square Kilometre Array<sup>7</sup> (SKA; Mellema et al. 2013; Koopmans et al. 2015) are seeking to detect the brightness temperature fluctuations in the cosmological 21-cm signal using statistical methods e.g. the power spectrum. Complementary to these 21-cm power spectrum measurement experiments, several efforts such as the Experiment to Detect the Global Epoch of Reionization Signature (EDGES; Bowman et al. 2018), the Large-aperture Experiment to Detect the Dark Ages (Bernardi et al. 2016), the Shaped Antenna measurement of the background Radio Spectrum 2 (Singh et al. 2017), the Sonda Cosmológica de las Islas para la Detección de Hidrógeno Neutro (Voytek et al. 2014), the Probing Radio Intensity at high  $z$  from Marion (Philip et al. 2018), and the Netherlands-China Low frequency Explorer<sup>8,9</sup> (NCLE) are seeking to measure the sky-averaged spectrum of the 21-cm signal.

At present, several instruments targeting the EoR redshifts have placed upper limits on the brightness temperature power spectrum of the redshifted 21-cm signal. Paciga et al. (2013) provided the first  $2\sigma$  upper limit on the brightness temperature of  $\Delta_{21}^2 < (248 \text{ mK})^2$  at wavenumber  $k \approx 0.5 \text{ h cMpc}^{-1}$  at redshift  $z = 8.6$  using the GMRT. Beardsley et al. (2016) used MWA to set a  $2\sigma$  upper limit of  $\Delta_{21}^2 < (164 \text{ mK})^2$  at  $k \approx 0.27 \text{ h cMpc}^{-1}$  at  $z = 7.1$ . The PAPER project also provided an upper limit of  $\Delta_{21}^2 < (22 \text{ mK})^2$  in the wavenumber range  $0.15 \leq k \leq 0.5 \text{ h cMpc}^{-1}$  at  $z = 8.4$  (Ali et al. 2015), but have recently retracted their claim due to issues with their analysis strategy (see the erratum Ali et al. 2018). The tightest  $2\sigma$  upper limit on the 21-cm power spectrum yet is  $\Delta_{21}^2 < (79.6 \text{ mK})^2$  at  $k \approx 0.053 \text{ h cMpc}^{-1}$  in the redshift range  $z = 9.6\text{--}10.6$  and was provided by Patil et al. (2017) using the LOFAR High Band Antenna (HBA) array. Instruments such as HERA, NenuFAR, and SKA-low which can potentially probe the CD redshifts are now in hardware roll-out stages (the latter is still in the development stage). Ewall-Wice et al. (2016) used low-frequency MWA observations (75–113 MHz) to place an upper limit of  $\Delta_{21}^2 < (10^4 \text{ mK})^2$  at  $k \approx 0.5$  on the power spectrum of the brightness temperature fluctuations of the 21-cm signal in the redshift range  $12 \lesssim z \lesssim 18$ , which in most models corresponds to the epoch of X-ray heating during the

CD (see e.g. Glover & Brand 2003; Fialkov & Barkana 2014; Ross et al. 2017).

In this work, we explore, for the first time, the possibility of observing the redshifted 21-cm signal from the CD era using the LOFAR-Low Band Antenna (LBA) array which observes in the 30–90 MHz frequency range. We use LOFAR-LBA dual-pointing observations of the North Celestial Pole (NCP field hereafter) and an adjacent field centred on the 3C220.3 radio galaxy (3C220 field hereafter), which is  $\sim 7^\circ$  away from the NCP, to study the challenges (systematic biases) in CD studies with the LOFAR-LBA and to set the first upper limits on the 21-cm brightness temperature power spectrum in the redshift range  $z = 19.8\text{--}25.2$ . We also demonstrate the application of a novel dual-pointing calibration strategy to calibrate the LOFAR-LBA data and the application of Gaussian Process Regression (GPR) as a powerful foreground removal technique in CD experiments.

The paper is organized as follows: in Section 2, we briefly describe the LOFAR-LBA system, the observational set-up, and pre-processing steps. In Section 3, we describe the multibeam calibration strategy to calibrate the LOFAR-LBA data. In Section 4, we assess the noise in the observed data and address the systematic biases, such as excess noise in Stokes  $I$  versus  $V$  using various statistical methods. We describe GPR in Section 5 and its application in removing residual foregrounds in LOFAR-LBA data. In Section 6, we determine the power spectra for both fields and derive upper limits on the 21-cm power spectrum in the redshift range  $z = 19.8\text{--}25.2$ . Finally, in Section 7, we summarize the work and discuss future prospects.

## 2 OBSERVATIONS AND PRE-PROCESSING

We used the LOFAR-LBA system with dual-pointing set-up to simultaneously observe the NCP field and the 3C220 field, which is  $\sim 7^\circ$  away from the NCP. The NCP is the primary target field of the LOFAR-EoR Key Science Project and has been used to set the first upper limits on the EoR power spectrum using LOFAR (see Patil et al. 2017). The observational set-up and preprocessing steps are described in the following subsections.

### 2.1 LOFAR-Low band array

The LOFAR-LBA consists of 38 stations spread across the Netherlands, providing shortest baseline lengths of  $\sim 80$  m and longest baseline lengths of  $\sim 100$  km. Out of these 38 stations, 24 stations (known as core stations) are spread within a core of 2 km radius, providing a densely sampled  $uv$ -plane. The remaining 14 stations (known as remote stations) are spread across the North-Eastern part of the Netherlands. Each LOFAR station consists of 96 low-band dual-polarization dipole antennas randomly spread within an area of 81 m diameter. The voltages measured with the cross dipoles are digitized using a 200 MHz sampling clock covering the frequency range of 10–90 MHz. The digitized data are beam-formed to produce a digitally steerable station beam. At a given time, only 48 out of 96 dipoles can be combined in the beam-former. This allows a user to choose from three different station configurations in LOFAR-LBA mode viz: LBA\_INNER where the 48 innermost dipoles (array width  $\sim 30$  m) are beam-formed, LBA\_OUTER where the 48 outermost dipoles (array width  $\sim 81$  m) are beam-formed, and LBA\_SPARSE where half of the innermost 48 dipoles, plus half of the outermost 48 dipoles (array width  $\sim 81$  m) are beam-formed. Each configuration results in a specific field of view (FoV) as well as different sensitivity due to mutual coupling between the dipoles. The

<sup>1</sup><http://www.lofar.org/>

<sup>2</sup><http://gmrt.ncra.tifr.res.in/>

<sup>3</sup><http://www.mwatelescope.org/>

<sup>4</sup><http://eor.berkeley.edu/>

<sup>5</sup><http://reionization.org/>

<sup>6</sup><https://nenufar.obs-nancay.fr/>

<sup>7</sup><http://skatelescope.org/>

<sup>8</sup><https://www.ru.nl/astrophysics/research/radboud-radio-lab-0/projects/net-herlands-china-low-frequency-explorer-ncle/>

<sup>9</sup><https://www.astron.nl/r-d-laboratory/ncle/netherlands-china-low-frequency-explorer-ncle>

**Table 1.** Observational details of the data.

Parameter	Value
Telescope	LOFAR LBA
Observation cycle and ID	Cycle 6, L557804
Antenna configuration	LBA_OUTER
Number of stations	38 (NL stations)
Observation start time (UTC)	2016 Nov 4; 16:21:44
Number of pointings	2
Phase centre ( $\alpha$ , $\delta$ ; J2000):	
NCP field	00h00m00s, +90° 00′ 00″
3C220 field	09h39m23s, +83° 15′ 26″
Duration of observation	14 h
Minimum frequency	38.08 MHz
Maximum frequency	85.54 MHz
Target bandwidth	48 MHz
Primary Beam FWHM	3.88° at 60 MHz
FoV	12 deg <sup>2</sup> at 60 MHz
SEFD	~25 kJy at 60 MHz
Polarization	Linear X-Y
Time, frequency resolution:	
Raw data	1 s, 3 kHz
After flagging step 1	2 s, 12 kHz (archived)
After flagging step 2	2 s, 61 kHz

LOFAR-LBA system has an instantaneous bandwidth of 96 MHz. However, multiple pointings in the sky can be traded against the observable bandwidth depending on the number of pointings. In the case of two pointings, the bandwidth is reduced to 48 MHz per pointing. Readers may refer to van Haarlem et al. (2013) for more information about the observation capabilities of LOFAR.

## 2.2 Observations

We use 14 h of synthesis observation data of the NCP and the 3C220 fields, which were observed simultaneously with dual beam pointings using the LBA\_OUTER mode of the LOFAR-LBA system. The data were recorded during LOFAR observation Cycle 6 (ID: L557804, 2016 November 4–5). The observational details of the data are summarized in Table 1. The digitized data from beam-formed stations were correlated with 1 s time resolution and 3 kHz frequency resolution. The recorded data consist of 244 sub-bands for each field within the frequency range of 38–86 MHz. Each sub-band has a width of 195.3 kHz and consists of 64 channels. The recorded correlations (XX, XY, YX, and YY) are stored in a measurement set (MS). The raw data volume for each field is ~18 TB and is pre-processed to reduce the data volume, as described in the next section.

## 2.3 Data selection and preprocessing

LOFAR-LBA has lower sensitivity and a relatively high Radio Frequency Interference (RFI) corruption level for frequencies above 70 MHz. Therefore, we used only 33 MHz bandwidth with the frequency range 39–72 MHz for pre-processing and further analysis. We used the standard LOFAR software pipeline (see e.g. LOFAR imaging cookbook<sup>10</sup>) for pre-processing the observed raw data. Processing steps include RFI-excision and averaging the data. Flagging of RFI corrupted data is performed on the highest resolution data

(1 s, 3 kHz) to minimize information loss. We use the AOFLAGGER software (Offringa et al. 2010; Offringa, van de Gronde & Roerdink 2012) to flag RFI corrupted data. Two channels on both edges of every sub-band were also discarded to avoid edge effects due to the polyphase filter. The remaining data were averaged in frequency and time to an intermediate resolution of 12 kHz and 2 s, resulting in 15 channels per sub-band. This intermediate-resolution data are archived for future use. To reduce the data volume further, it was averaged in frequency to 61 kHz and the autocorrelations were also flagged. The resulting data consist of three channels of 61 kHz each per sub-band and has a time resolution of 2 s. We flagged the remote station RS503LBA in all sub-bands for both fields because of its proximity to a windmill, which causes strong RFI in the visibilities of the station. We also observed that CS302LBA had poor gain upon inspecting the visibilities and flagged it for both fields. The flagging and averaging were performed separately on both 3C220 and NCP field datasets, although some correlation is obviously expected. Fig. 1 shows LOFAR-LBA *uv*-coverage (the inner region with  $|u| < 600\lambda$ ) for the 3C220 field pointing for 14 h track at 60 MHz after exclusion of flagged visibilities and the radial profile of the *uv*-coverage.

## 3 CALIBRATION SCHEME

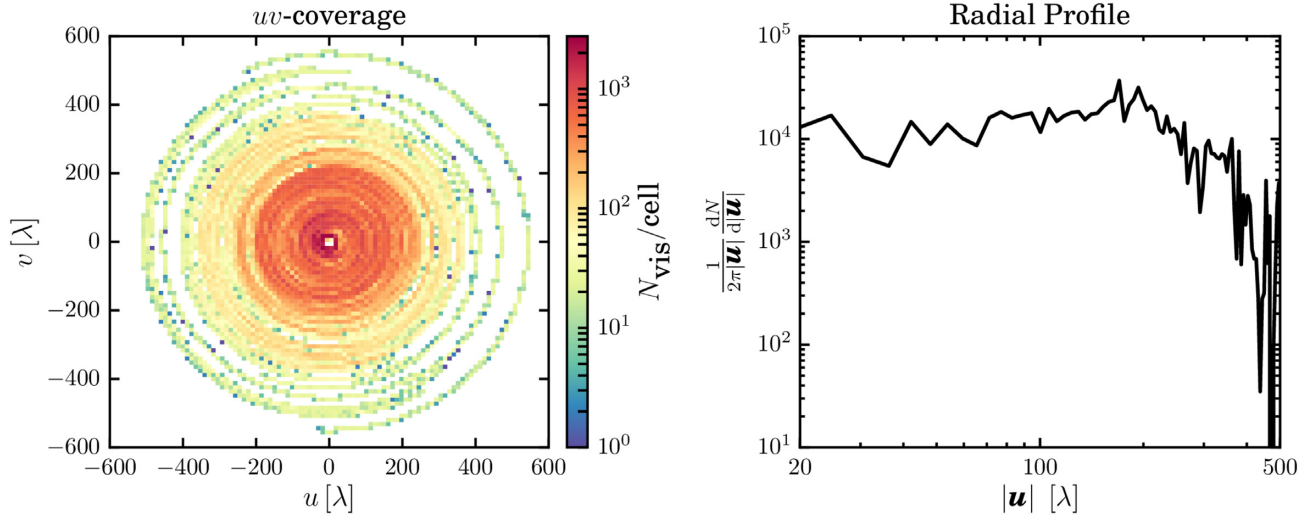
The visibilities recorded by LOFAR are corrupted by the instrumental (complex station gains, primary beam, instrumental bandpass, clock drift, etc.) and environmental (ionosphere) factors. Calibration of the LOFAR-LBA system involves estimating the errors that corrupt the measured visibilities, and to obtain an accurate estimate of the true visibilities from observed data. Calibration of LOFAR-LBA data involves two major steps: (a) direction independent (DI) calibration and, (b) direction dependent (DD) calibration. DI calibration involves estimation of a single instrumental gain (represented by a complex  $2 \times 2$  Jones matrix) for each beam-formed station, and DD calibration accounts for the direction dependent errors arising from wave propagation effects through the ionosphere and the primary beam. We use SAGECAL-CO<sup>11</sup> to perform the major calibration steps. SAGECAL-CO performs calibration in a distributed way using consensus optimization (Boyd et al. 2011), which is an effective way to improve the quality of the calibration of radio interferometric data. In SAGECAL-CO, the calibration problem is transformed into consensus optimization by adding frequency smoothness of systematic errors as a constraint. It uses an Alternating Direction Method of Multipliers (ADMM) algorithm to reach convergence. Readers may refer to Yatawatta (2015, 2016, 2018), Yatawatta et al. (2017) for a detailed description of the SAGECAL-CO algorithm and its capabilities.

Upon inspection of the raw visibilities, we observed that Cas A (~30° away from NCP) and Cyg A (~50° away from the NCP) superpose significant side-lobes onto both fields. It is crucial to subtract these sources before performing DI calibration to avoid errors due to these side-lobes. We use DD calibration in SAGECAL-CO to subtract Cas A and Cyg A. Gehlot et al. (2018, G18 hereafter) showed that the residuals after subtraction of bright sources such as Cas A and Cyg A are significant as well as incoherent over time-scales of a few minutes, depending on the strength of ionospheric scintillations. Therefore, we use a solution time and frequency interval of 30 and 61 kHz to subtract Cas A and Cyg A, which is optimized to incorporate ionospheric effects while maintaining a

<sup>10</sup><https://www.astron.nl/radio-observatory/lofar/lofar-imaging-cookbook>

<sup>11</sup><http://sagecal.sourceforge.net/>





**Figure 1.** Left-hand panel: the inner region ( $|u| < 600\lambda$ ) of LOFAR-LBA  $uv$ -coverage for the 3C220 field for 14 h track at 60 MHz after excluding flagged visibilities. Right-hand panel: the radial  $uv$ -density  $\frac{dN}{2\pi|u|d|u|}$  corresponding to the left-hand panel.

decent signal-to-noise ratio ( $\gtrsim 10$ ) for the given solution interval. We use the Cas A and Cyg A shapelet models<sup>12</sup> as an input model for calibration and subtraction. The subtraction was performed individually on both fields.

The two fields, 3C220 and NCP, given their different pointings and gain solutions, have slightly different morphologies. The 3C220 field consists of a reasonably bright source located at the phase centre (the 3C220.3 radio galaxy with a flux of  $\sim 38$  Jy at 74 MHz, Cohen et al. 2007) which can be utilized as a bandpass calibrator, making calibration of the 3C220 field fairly straightforward. However, the NCP field does not have such relatively bright sources near the phase centre, which makes it more difficult to calibrate the field. Therefore, we adopt a calibration strategy where we calibrate the 3C220 field first and then use the output calibration products to calibrate the NCP field, given that the bandpass calibration solutions should be similar between the fields because of the same electronics, and that any effect of the beam should be spectrally smooth near the phase centre. A similar technique to calibrate the LOFAR-LBA data to study the ionospheric effects is shown in de Gasperin et al. (2018, 2019). Similar types of calibration strategies are more common in radio survey experiments, although in those cases it is often required to switch between sources in time. The calibration settings (e.g. solution interval, frequency resolution, ADMM iterations, and regularization factor) for the two fields are chosen to account for any rapidly varying effects in time and frequency such as the ionosphere while maintaining a reasonable signal-to-noise ratio per solution interval. Most of these settings are decided on the basis of the analysis and lessons learned in G18 as well as the analysis of the LOFAR-EoR data (see e.g. Patil et al. 2017).

### 3.1 Calibrating the 3C220 field

To calibrate the 3C220 field, we use a calibration strategy similar to that discussed in G18. The 3C220.3 radio galaxy is a double-lobed source of  $\sim 8$  arcsec extent, but it is unresolved with the

LOFAR-LBA array which has a maximum resolution of  $\sim 15$  arcsec. Therefore, we use a single point source representing 3C220.3 with 38 Jy flux at 74 MHz (Cohen et al. 2007) and a spectral index of  $-0.8$  as a starting model for DI calibration. The major steps involved in the calibration of the 3C220 field are as follows:

(i) Calibrate the raw visibilities using the 3C220.3 point source model in NDPPP<sup>13</sup> to obtain the station gain solutions with 30 s and 61 kHz calibration solution intervals and subsequently apply them to the data. This step is performed separately for each sub-band (without consensus optimization). We include the primary beam<sup>14</sup> in the calibration step in NDPPP. Note that the LOFAR-LBA beam model has only been implemented in NDPPP at present. Hence, it is utilized for primary DI calibration for both fields. Note that we do not exclude any baselines during DI calibration steps for both fields.

(ii) Deconvolve (clean) and image the calibrated visibilities using the WSCLEAN package (Offringa et al. 2014) with the following settings: cleaning threshold =  $0.5\sigma$ , weighting scheme = uniform, imaging baseline range =  $0-5000\lambda$ , fourth-order linear polynomial<sup>15</sup> for fitting the source spectrum over 15 points which correspond to averaged flux over 2.2 MHz bands spread within 33 MHz bandwidth. The cleaning parameters are chosen such that the modelled sources with lowest flux values are still a factor of few above the confusion limit at 60 MHz ( $\sigma_c \sim 10$  mJy beam $^{-1}$ , see van Haarlem et al. 2013 for calculation of  $\sigma_c$ ). Since we do not apply the primary beam correction during imaging, the source fluxes are apparent and their spectra also possess the primary beam variations which are less smooth compared to the intrinsic

<sup>13</sup>[http://www.lofar.org/operations/doku.php?id=public:user\\_software:ndppp](http://www.lofar.org/operations/doku.php?id=public:user_software:ndppp)

<sup>14</sup>Current LBA primary beam models are based on electromagnetic (EM) simulations of the LOFAR-LBA dipoles (private communication with LOFAR Radio Observatory).

<sup>15</sup>Using log polynomials to fit source spectra is unstable in WSCLEAN. Therefore, we use an ordinary fourth-order linear polynomial to fit source spectra. However, SAGECAL-CO is only compatible with log polynomials. Therefore, we separately fit the source spectra with a third-order log-polynomial to make it compatible with SAGECAL-CO.

<sup>12</sup>Cas A and Cyg A models were derived from wide-band LOFAR-LBA and HBA observations of Cas A and Cyg A. Each source has about 200 components (shapelets and point). See Yatawatta(2011) for more details.

source spectra. Using a fourth-order polynomial for spectral fitting easily captures these beam variations compared to a lower order polynomial and facilitates better source subtraction. Step (i) is repeated once more using the clean model of 3C220.3 obtained in step (ii) and deconvolution is performed to obtain a more accurate 3C220.3 clean model. Further iterations were not required as the model converged.

(iii) Use SAGECAL-CO to perform DI calibration of raw visibilities and subtract 3C220.3 using consensus optimization (seven iterations and regularization factor of 5) over a 33 MHz frequency range. We provide the final clean model of 3C220.3 obtained after step (ii) as input to SAGECAL-CO and use a calibration solution interval of 30 s and 183.1 kHz. The obtained gain solutions are subsequently applied to the residual visibilities.

(iv) Repeat the deconvolution with the same settings (but with lower clean mask =  $4\sigma$ ) in step (ii) to clean and image the residual visibilities after step (iii). The output clean model of the radio sources in the field contains 1270 components (points plus Gaussians) with flux  $>40$  mJy at 55 MHz. We repeated step (iii) with this updated sky model to perform DI calibration and subtraction of 3C220.3 from the visibilities. Using a more complete sky-model in DI calibration allows for the mitigation of calibration errors due to unmodelled sources and produces accurately calibrated visibilities. The gain solutions obtained after this step are later utilized in the calibration of the NCP field.

(v) Use DD calibration with SAGECAL-CO to subtract the clean model obtained in step (iv). SAGECAL-CO accounts for DD errors by obtaining the gain solutions in multiple directions. It subtracts the sources in each direction by multiplying the obtained gain solutions with the predicted visibilities and subtracting the product from the observed visibilities. We divide the 1270 components into four clusters using the weighted  $K$ -means clustering algorithm (Kazemi, Yatawatta & Zaroubi 2013) and use the centres of these clusters as four different directions. This algorithm operates on a unit sphere and the corresponding weights are determined by the source flux. The algorithm creates clusters such that the cluster size is minimized while maintaining similar net flux in each cluster to ensure sufficient signal-to-noise ratio for each cluster. We use a gain solution interval of 20 min and 183.1 kHz and 20 ADMM iterations for each gain solution while keeping the same regularization factor of  $\rho = 5$  (Yatawatta 2016) as in DI calibration. We discard the baselines  $\leq 200\lambda$  in the DD calibration. These excluded baselines ( $<200\lambda$ ) are used for further analyses and power spectrum estimation. Using a baseline cut avoids any bias due to unmodelled diffuse emission on shorter baselines excluded from calibration (see e.g. Patil et al. 2016; Barry et al. 2016; Ewall-Wice et al. 2017; Gehlot et al. 2018 for more details). It also mitigates the suppression of the 21-cm signal caused by the use of an incomplete sky model in the calibration, as shown in Patil et al. (2017) and Mouri Sardarabadi & Koopmans (2018), and we will test this further in future. However, the exclusion of short baselines from the calibration also has a drawback that longer baselines are prone to calibration errors. These errors, when applied to excluded baselines, cause the foregrounds to leak into the ‘EoR-window’ on corresponding baselines (Barry et al. 2016; Patil et al. 2016). Using the smoothness of gain solutions as a constraint in the calibration largely mitigates this effect (Mouri Sardarabadi & Koopmans 2018). An optimal baseline selection criteria for calibration which accounts for these effects itself requires fairly rigorous analysis and is beyond the scope of this paper. The choice of  $200\lambda$  baseline cut comes from the reason that the radial profile of the  $uv$ -coverage (see right-hand panel of Fig. 1) is relatively flat within  $20\lambda \leq |u| \leq 200\lambda$  range and drops for longer

baselines. This is an optimal choice for power spectrum estimation because of relatively uniform  $uv$ -coverage.

(vi) Image the residual visibilities in step (v) with WSCLEAN. We used the following settings: weighting scheme = natural, pixel size = 3 arcmin, image dimensions =  $300 \times 300$  pixels, and imaging baselines =  $15\text{--}200\lambda$ . Note that we do not deconvolve the final residual images. The output Stokes  $I$ ,  $V$ , and point spread function (PSF) images were stored for further analysis. The left-hand panel of Fig. 2 shows the dirty continuum image of the 3C220 field after DI calibration where the 3C220.3 has been subtracted.

### 3.2 Calibrating the NCP field

The absence of very bright sources makes the NCP field more difficult to calibrate using the strategy we employed for the 3C220 field. Therefore, we utilize a different approach. The NCP field consists of a moderately bright source (3C061.1) which lies at the edge of the primary beam causing the source to exhibit peculiar behaviour in its gain solutions. We, therefore, subtract 3C061.1 from the raw visibilities using DD calibration with SAGECAL-CO with the same settings as we employed for the Cas A and Cyg A subtractions. The 3C061.1 input model is adapted from the intrinsic model of 3C061.1 (points + shapelets, at 150 MHz) used in the LOFAR-EoR data processing pipeline (see e.g. Patil et al. 2017). The fluxes in the model were scaled properly to match the flux values quoted in Laing & Peacock (1980) and Hales et al. (1995). After subtraction of 3C061.1, visibilities were calibrated using the following steps:

(i) Apply the DI gain solution amplitudes of the 3C220 field obtained in step (iv) in Section 3.1 to the NCP field visibilities to set the amplitude scale.

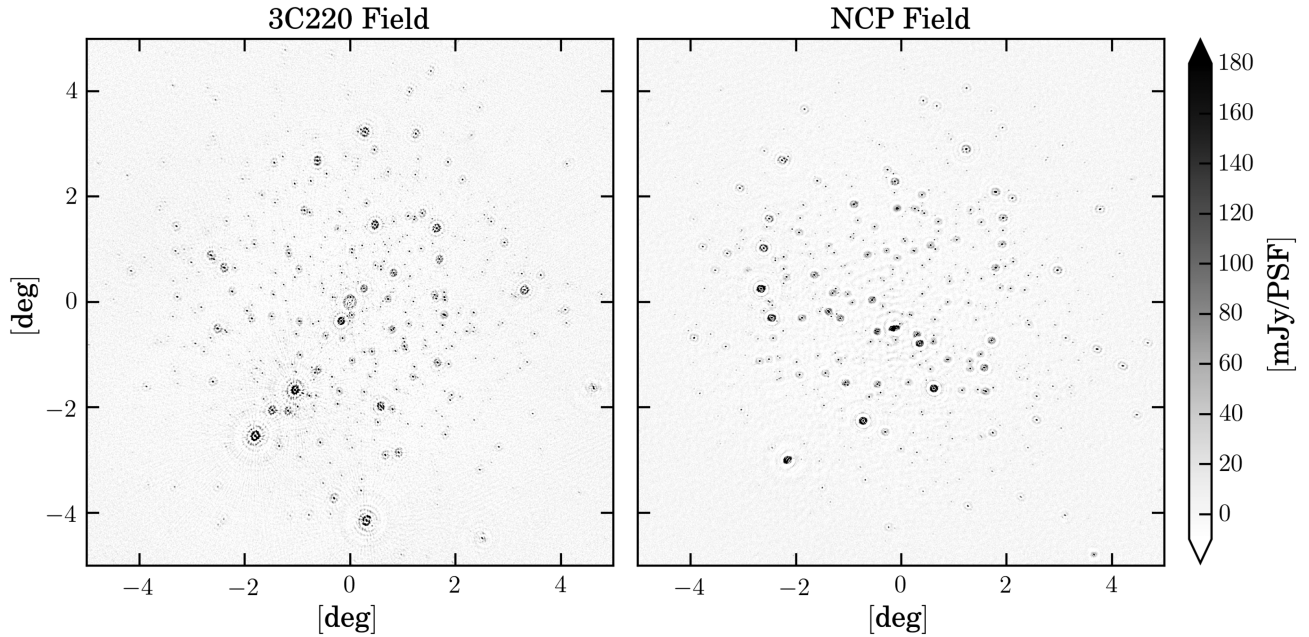
(ii) Deconvolve (clean) and image the resulting visibilities using WSCLEAN with the following settings: cleaning threshold =  $0.5\sigma$ , weighting scheme = uniform, imaging baseline range =  $0\text{--}2000\lambda$ , and second-order polynomial for fitting the source spectrum over five points which correspond to an averaged flux over 6.6 MHz bands spread over 33 MHz.

(iii) Perform DI calibration of the visibilities with SAGECAL-CO using consensus optimization (with same settings as in DI calibration of the 3C220 field) over the 33 MHz frequency range. The clean model obtained in step (ii) is provided as input. We use a calibration solution interval of 10 min and 183.1 kHz. The obtained gain solutions are subsequently applied to the visibilities. We repeat steps (ii) and (iii) in a self-cal loop with three iterations. The final clean model after 3 self-cal iterations contains 1470 components (points plus Gaussians) with flux  $>40$  mJy at 55 MHz.

(iv) Perform phase calibration using NDPPP on the visibilities obtained after step (i). We use the final clean model obtained after step (iii) as input and choose 30 s, 183.1 kHz as the calibration solution interval.

(v) Use DD calibration with SAGECAL-CO to subtract the clean-model obtained in step (iii). We divide 1470 components into three clusters representing three directions (which represent three non-overlapping regions within the primary beam) using the weighted  $K$ -means clustering algorithm (same as in step (v) of Section 3.1). We use a gain solution interval of 20 min and 183.1 kHz and 20 ADMM iterations for each gain solution. We discard the baselines  $\leq 200\lambda$  to avoid errors due to unmodeled diffuse emission on shorter baselines and to avoid signal suppression.

(vi) Image the residual visibilities in step (v) with WSCLEAN using the following settings: weighting scheme = natural, pixel size



**Figure 2.** Left- and right-hand panels show Stokes  $I$  continuum ‘dirty’ images (39–72 MHz) of the 3C220 and NCP field respectively, after DI calibration. The images are not cleaned, and were produced using  $\leq 2000\lambda$  baselines with the Briggs  $-0.1$  weighting scheme. The observed image rms is  $\sigma_{\text{rms}} \sim 7$  mJy for the 3C220 field and  $\sim 5.5$  mJy for the NCP field, respectively. These values are still  $\sim 10$  times higher than the expected rms ( $\sim 0.7$  mJy) calculated using SEFD estimates for LOFAR-LBA. The values of  $\sigma_{\text{rms}}$  can be calculated from SEFD using the relation:  $\sigma_{\text{rms}} = \text{SEFD} / \sqrt{2N(N-1)\Delta\nu\Delta t}$ , where SEFD  $\sim 30$  kJy at 55 MHz,  $N = 29$  (corresponding to  $2000\lambda$  baseline range),  $\Delta\nu = 33$  MHz, and  $\Delta t = 0.9 \times 14$  h (assuming flagged data at 10 per cent level).

$= 3$  arcmin, image dimensions  $= 300 \times 300$  pixels, and imaging baselines  $= 15\text{--}200\lambda$ . The output Stokes  $I$ ,  $V$ , and PSF images were stored for further analysis. The right-hand panel of Fig. 2 shows the dirty continuum image of the NCP field after DI calibration.

We only use the beam model during the DI calibration and image deconvolution steps, and we do not correct the residual images for the primary beam. Also, we do not analyse the Stokes  $Q$  and  $U$  data. The latter is mainly because we do not include any polarized (compact or diffuse) emission in sky model used for the calibration. Any unmodeled emission in Stokes  $Q$  and  $U$  can essentially bias the calibration. The currently utilized calibration scheme is defined such that only the Stokes  $I$  and  $V$  are constrained by the sky model, whereas, the Stokes  $Q$ , and  $U$  have the freedom to rotate. Moreover, in the rotation measure (RM)-synthesis analysis in G18, we did not observe any signature of the polarized emission in RM-space, suggesting the absence of significant polarized emission at these low frequencies. Because the RM scales as  $\lambda^2$ , any polarized emission at low frequencies (40–70 MHz) is essentially depolarized by the intervening magneto-ionic medium and the rapidly varying ionosphere. However, in G18, we observed strong polarization leakage of the bright off-axis source Cas A from Stokes  $I$  to  $Q$ ,  $U$ , and also in Stokes  $V$  but at a much weaker level compared to Stokes  $Q$  and  $U$ . This effect is mitigated by subtraction of Cas A and Cyg A during using DD calibration at higher time resolution (30 s) and is already accounted for in the current analysis. The leakage of (partly) polarized foregrounds to Stokes  $I$ , if any, is expected to be significantly lower than the current noise level and currently does not affect our analysis.

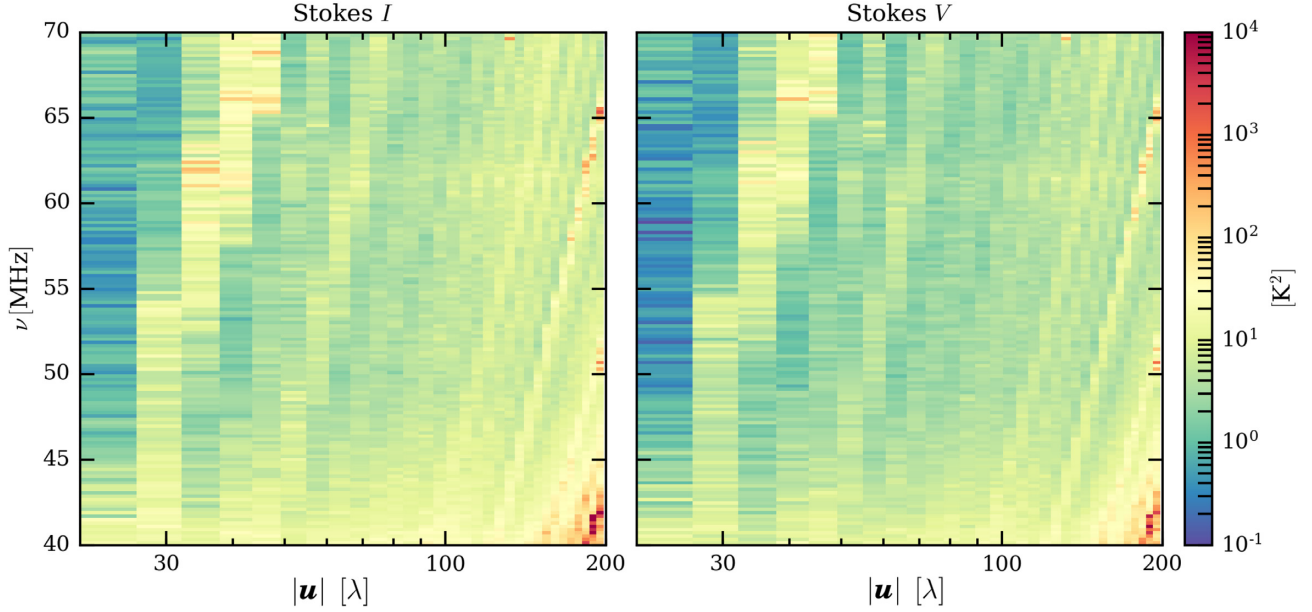
At this point, we have residual data cubes that are DI calibrated and where the sky model has been subtracted using their DD gain solutions. These residual cubes form the input for subsequent analyses. In the following sections, we will discuss these analyses.

#### 4 NOISE STATISTICS IN LOFAR-LBA

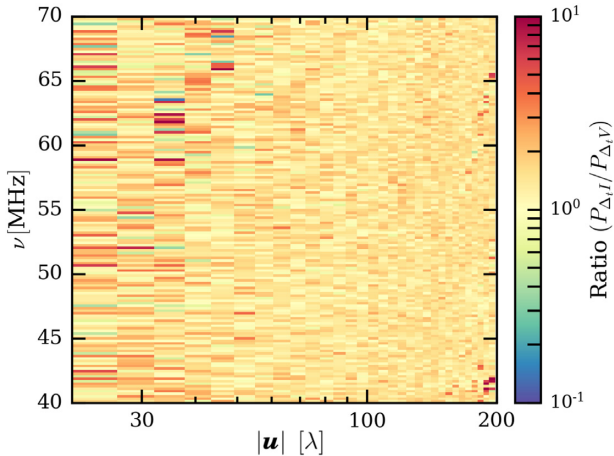
Current estimates of the average signal equivalent flux density (SEFD) per station of the LOFAR-LBA array are derived from the observations of bright sources at zenith. However, the SEFD of LOFAR varies as a function of angle from the zenith. Therefore, using zenith SEFD estimates to derive the noise on the visibilities and rms in the images (also noise power spectra) typically underestimates the SEFD for the fields away from the zenith. To avoid this bias, we estimate the noise and hence the noise spectrum (in baseline-frequency space) for the 3C220 field from the visibilities. A standard method to estimate the noise on visibilities is to subtract the ungridded visibilities corresponding for two contiguous time-steps at the highest time resolution. However, this method is not feasible for large LOFAR-LBA data sets ( $\sim 18$  TB per data set) because of a large number of baselines and time-steps. Therefore, we use an alternative approach where we estimate the noise spectrum from the gridded visibilities (see e.g. Jacobs et al. 2016; Beardsley et al. 2016; Ewall-Wice et al. 2016). We expect that these two approaches become equivalent to each other for data sets with a large number of time-samples and leave additional comparison tests between the two approaches for future analyses.

We split the DI-calibrated visibilities of the 3C220 field into even and odd samplings with 12 s cadence such that these samplings are interleaved in time. Note that for the baseline range  $20\lambda \leq |u| \leq 200\lambda$  which we probe in our analysis, the sky and the PSF do not vary over a 12 s interval. Also, any sky leakage over 12 s will appear as a wedge in the cylindrically averaged power spectrum, which we do not observe in the analyses (shown in later sections). Moreover, we expect the system to be coherent over 12 s and only ionospheric effects are expected to change. We image these even and odd samplings using WSCLEAN with the ‘natural’ weighting scheme. We Fourier transform (FT) the even and odd image cubes and properly scale visibilities in each  $uv$ -cell with corresponding sampling





**Figure 3.** Stokes  $I$  ( $P_{\Delta I}(|\mathbf{u}|, \nu)$ ) and  $V$  ( $P_{\Delta V}(|\mathbf{u}|, \nu)$ ) noise spectra for the 3C220 field. Left- and right-hand panels correspond to Stokes  $I$  and  $V$ , respectively.



**Figure 4.** The ratio  $P_{\Delta I}/P_{\Delta V}$  of the noise spectra shown in Fig. 3. The ratio is flat except for a few outliers at shorter baselines.

density to remove the effect of gridding weights during imaging. We calculate the azimuthally averaged (spatial) power spectrum of the difference as  $P_{\Delta I}(|\mathbf{u}|, \nu) \equiv \langle \Delta_t \tilde{I} \rangle^2 = \langle \tilde{I}_{\text{even}} - \tilde{I}_{\text{odd}} \rangle^2 / 2$ , where  $\tilde{I}_{\text{even}}$  and  $\tilde{I}_{\text{odd}}$  are the FT of the even and odd Stokes  $I$  image cubes, respectively,  $\mathbf{u} = (u, v)$  is the baseline vector (in units of wavelength) in the  $uv$ -plane and  $|\mathbf{u}| = \sqrt{u^2 + v^2}$  and  $\nu$  is the frequency. Similarly,  $P_{\Delta V}(|\mathbf{u}|, \nu) \equiv \langle \Delta_t \tilde{V} \rangle^2$  is determined using the even and odd Stokes  $V$  image cubes.

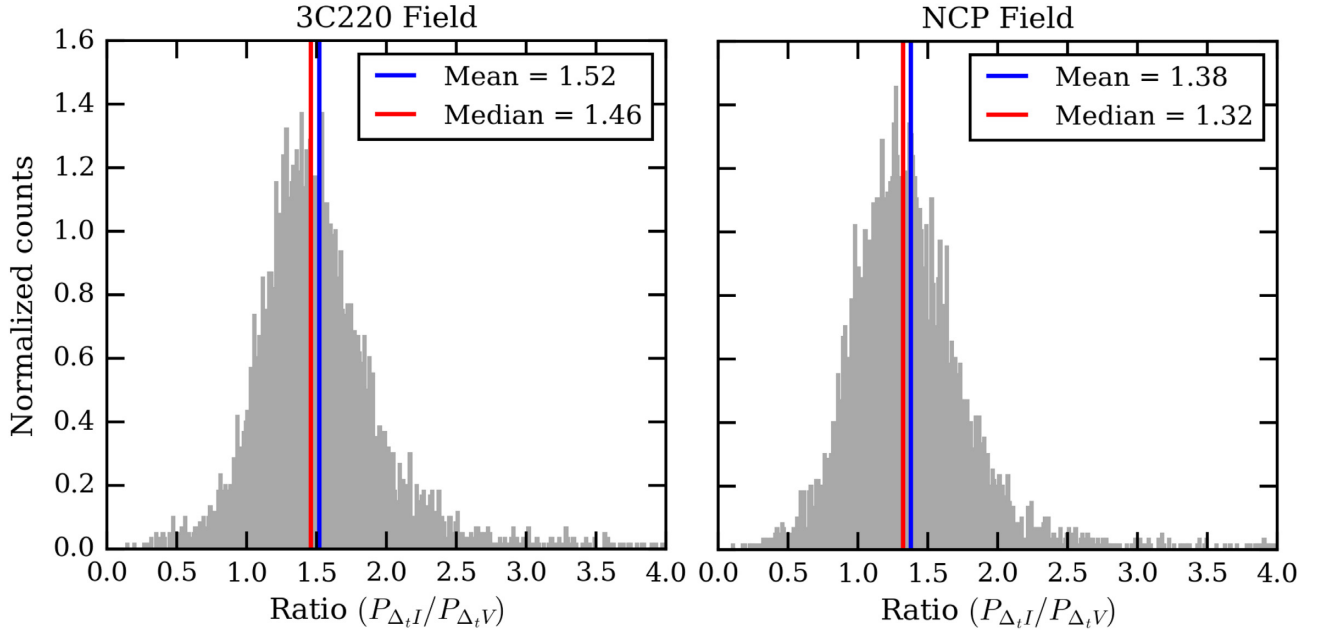
#### 4.1 Physical excess noise

Fig. 3 shows  $P_{\Delta I}$  and  $P_{\Delta V}$  for the 20–200 $\lambda$  baseline range for the 3C220 field. We observe that both  $P_{\Delta I}$  and  $P_{\Delta V}$  spectra are relatively flat. The bright tilted streaks are a consequence of varying  $uv$ -density as a function of baseline length in LOFAR-LBA. We compare  $P_{\Delta I}$  and  $P_{\Delta V}$  by calculating their ratio. Fig. 4 shows the ratio  $P_{\Delta I}/P_{\Delta V}$  of the spectra shown in Fig. 3. We observe that the ratio is remarkably flat, except for a few outliers at shorter

baselines ( $\leq 40\lambda$ ). Most of these outliers are also coincident with baselines where  $uv$ -density is relatively low. These outliers might arise due to imperfect calibration and slight differences in flagging of RFI affected baselines post calibration along with  $uv$ -density variations. Ideally, if the noise properties of Stokes  $I$  and  $V$  are statistically identical and if the sky and the PSF do not change over a 12 s interval,  $P_{\Delta I}$  and  $P_{\Delta V}$  are expected to be identical assuming that the sky has a negligible circular polarized emission component and Stokes  $V$  is virtually empty. However, we observe excess power in Stokes  $I$  compared to Stokes  $V$ , which is largely constant over the 20–200 $\lambda$  baseline range and over the 30 MHz bandwidth. Although the power in both Stokes  $I$  and  $V$  varies slightly with increasing baseline length, the ratio remains constant, suggesting that this slight variation is a result of varying  $uv$ -density. Most correlations that are spectrally smooth, e.g. due to intrinsic foregrounds, instrumental mode-mixing and ionosphere, appear as a ‘wedge’ in the 2D power spectrum. Whereas systematic effects with specific spectral structure e.g. cable reflections may appear as a distinct feature above the ‘wedge’. However, only those effects that are non-smooth in frequency or possess noise-like behaviour affect most scales in the 2D power spectrum. In later sections (see Section 6.3) we show that the corresponding 2D power spectra for Stokes  $I$  and  $V$  noise estimates are featureless and devoid of any ‘wedge’ like structure or other distinct features corresponding to systematic effects. Hence this physical excess noise, for all practical purposes, is treated as additional white noise in Stokes  $I$  that is seemingly uncorrelated in frequency and remains more or less the same for different baseline lengths.

The left-hand panel of Fig. 5 shows the normalized histogram of the distribution of  $P_{\Delta I}/P_{\Delta V}$  values for the 3C220 field. The distribution has a median value of 1.46 and a mean value of 1.54, with most values lying within the range 1–2. The noise spectra and their ratio for the NCP field also exhibit similar behaviour as the 3C220 field that the ratio  $P_{\Delta I}/P_{\Delta V}$  is flat in frequency-baseline space. However, the distribution of the ratio values (see right-hand panel of Fig. 5) has a slightly lower median and mean values of 1.32 and 1.38, respectively. The cause of this excess power in  $P_{\Delta I}$  is still unknown, but it is higher for the 3C220 field which has a





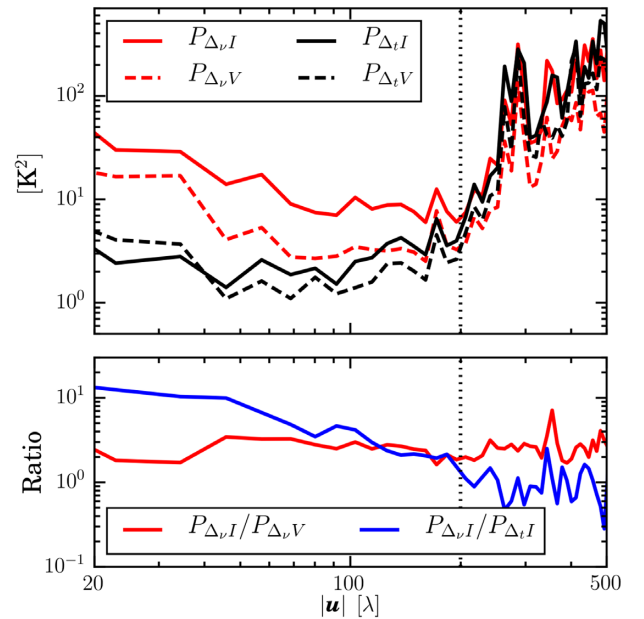
**Figure 5.** The left- and the right-hand panels show normalized histograms of the distribution of the ratio values  $P_{\Delta_t I}/P_{\Delta_t V}$  for the 3C220 and the NCP fields, respectively. The red and blue vertical lines represent the median and the mean of the distribution, respectively. For the 3C220 field, the distribution has a median value of 1.46 and a mean value of 1.52. Similarly, the median and the mean values for the NCP field are 1.32 and 1.38, respectively.

bright source at the centre, compared to the NCP field which is devoid of relatively bright sources. We are currently investigating the cause of this excess, but given that the excess is different for the two fields, ionospheric or interplanetary scintillation noise might be a probable reason for this excess, although the rapid decorrelation with frequency remains unexplained.

#### 4.2 Variance at the intersub-band level

We use the azimuthally averaged power spectrum of the difference of Stokes  $I$  and  $V$  images between two contiguous sub-bands (differential power spectrum) to study the behaviour of variance at the intersub-band level (see e.g. Patil et al. 2016; Gehlot et al. 2018). This method is based on the assumption that Stokes  $I$  images are composed of total sky signal convolved with the PSF plus additive noise. Assuming that the sky signal, which is smooth in frequency does not change<sup>16</sup> between two consecutive sub-bands 195 kHz apart, and any contribution due to the sky signal should largely drop out in the differential Stokes  $I$  images. Similarly, differential Stokes  $V$  images should contain only noise. However, effects which are non-smooth at the sub-band level are expected to leave their signature in the differential Stokes images.

We use Stokes  $I$  and  $V$  residual images of the 3C220 field ( $\nu_1 = 59.76$  MHz and  $\nu_2 = 59.95$  MHz, located at the most sensitive part of the band) after DD calibration to estimate the differential power spectra  $P_{\Delta_\nu I}$  and  $P_{\Delta_\nu V}$ , and determine their ratio  $P_{\Delta_\nu I}/P_{\Delta_\nu V}$ . The top panel of Fig. 6 shows  $P_{\Delta_\nu I}$  (red solid curve) and  $P_{\Delta_\nu V}$  (red dashed curve). We also show a slice of  $P_{\Delta_t I}$  and  $P_{\Delta_t V}$  at 59.95 MHz in the same plot for comparison. We observe that the power spectra are more or less flat on baselines  $|\mathbf{u}| > 200\lambda$  and increase rapidly for



**Figure 6.** Top panel: the differential Stokes  $I$  and  $V$  power spectra calculated using residual images of the 3C220 field. The solid red curve corresponds to  $P_{\Delta_\nu I}$  and the dashed red curve corresponds to  $P_{\Delta_\nu V}$ . The solid and dashed black curves correspond to  $P_{\Delta_t I}$  and  $P_{\Delta_t V}$ , respectively, at  $\nu = 59.95$  MHz. Bottom panel: the ratio  $P_{\Delta_\nu I}/P_{\Delta_\nu V}$  (red curve) and the ratio  $P_{\Delta_\nu I}/P_{\Delta_t I}$  (blue curve). The dotted vertical line shows the location of the  $200\lambda$  baseline cut.

$|\mathbf{u}| > 200\lambda$ . This can be attributed to variations in the  $uv$ -coverage of LOFAR-LBA. The variations in these power spectra also correlate well with the  $uv$ -coverage profile shown in Fig. 1.

The bottom panel of Fig. 6 shows the ratio  $P_{\Delta_\nu I}/P_{\Delta_\nu V}$  (red curve) and the ratio  $P_{\Delta_\nu I}/P_{\Delta_t I}$  (blue curve). We also observe that the ratio

<sup>16</sup>For a spectral index of  $-2.55$ , sky brightness changes at  $\sim 0.8$  per cent level for 195 kHz frequency separation at 60 MHz, which has a negligible contribution to the difference.

$P_{\Delta_v I}/P_{\Delta_v V}$  is relatively flat and has value  $\sim 2-3$  over the presented baseline range. This suggests that the rapid upturn in the power spectra shown in the top panel is due to  $uv$ -coverage variations and cancels out in the ratio. The excess variance in  $P_{\Delta_v I}$  compared to  $P_{\Delta_v V}$  is possibly due to random errors produced in calibration and/or erratic ionosphere. These random errors when applied to the data or the sky model during subtraction, affect both Stokes  $I$  and  $V$ . However, these errors lead to larger variance when applied to the emission in Stokes  $I$  compared Stokes  $V$  which lacks any emission (or below thermal noise, if any), resulting in excess noise at sub-band level.

The ratio we observe here is considerably smaller than that we observed in G18 ( $P_{\Delta_v I}/P_{\Delta_v V} \gtrsim 10$ ). This lower Stokes  $I$  noise is in part achieved because SAGECAL-CO enforces frequency smoothness of the gain solutions in the calibration process, and also because the ionospheric activity is more benign compared to the observation in G18 where frequency smoothness was not enforced in the calibration and the ionosphere behaved erratically. To quantify the ionospheric activity, we use ionospheric RM estimates from the GPS data. The ionospheric RM levels for the current observation are of order  $\sim 0.1 - 0.15 \text{ rad m}^{-2}$  during 90 per cent of the observation which is  $\gtrsim 4$  times lower than the ionospheric RM levels ( $\text{RM} > 0.4 \text{ rad m}^{-2}$ ) for the observation in G18, suggesting benign ionospheric activity.

Furthermore, from comparison of  $P_{\Delta_v I}$  with  $P_{\Delta_t I}$ , we observe that there is a sudden jump in the ratio at  $|u| \sim 200\lambda$ . The ratio is  $\gtrsim 2$  for  $|u| < 200$  and it continues to increase as the baseline length decreases, compared to the values ( $\sim 1-2$ ) for  $|u| > 200$ . We attribute this effect to the  $200\lambda$  baseline cut used in the DD calibration. The sky-model incompleteness or ionospheric effects can introduce random errors during the calibration step. These random errors on gain solutions when applied to the baselines excluded during the calibration step, increase the variance in Stokes  $I$  compared to Stokes  $V$  on excluded baselines (Patil et al. 2016; Barry et al. 2016; Ewall-Wice et al. 2016, 2017; Gehlot et al. 2018; Mouri Sardarabadi & Koopmans 2018).

## 5 GAUSSIAN PROCESS REGRESSION

After subtracting the calibration sky model using DD calibration, any remaining foreground emission within the primary beam consists of unresolved sources, sources below the confusion noise, sources not included in the model, and diffuse emission. These foregrounds should vary slowly with frequency, making them separable from the 21-cm signal which has rapid spectral fluctuations. We use a GPR technique (see Mertens, Ghosh & Koopmans 2018 for more details) to remove this remaining foreground emission and other spectral structures imparted on the data due to instrumental mode-mixing, such as instrumental chromaticity and imperfect calibration residuals. In this section, we briefly describe GPR and its application to LOFAR-LBA data.

### 5.1 Methodology

The visibilities observed by an interferometer ( $\mathcal{V}_{\text{obs}}(\mathbf{u}, \nu)$ ) can be written as a sum of different components viz. the signal of interest ( $\mathcal{V}_{21}(\mathbf{u}, \nu)$ ), the foreground contribution ( $\mathcal{V}_{\text{sky}}(\mathbf{u}, \nu)$ ), instrumental mode-mixing ( $\mathcal{V}_{\text{mix}}(\mathbf{u}, \nu)$ ), and noise ( $\mathcal{V}_n(\mathbf{u}, \nu)$ ), i.e.

$$\mathcal{V}_{\text{obs}}(\mathbf{u}, \nu) = \mathcal{V}_{21}(\mathbf{u}, \nu) + \mathcal{V}_{\text{sky}}(\mathbf{u}, \nu) + \mathcal{V}_{\text{mix}}(\mathbf{u}, \nu) + \mathcal{V}_n(\mathbf{u}, \nu). \quad (1)$$

Each of these components has a distinct spectral behaviour which is exploited by GPR to separate them from each other and eventually

remove the foreground components from the observed visibilities, leaving residuals with the signal of interest buried below the noise (Mertens et al. 2018). GPR models these different components with Gaussian Processes (GP). A GP ( $f \sim \mathcal{GP}(m, \kappa)$ ) is the joint distribution of a collection of normally distributed random variables and is defined by its mean  $m$  and covariance function  $\kappa$ . Values of  $\kappa$  specify the covariance between pairs of points at different frequencies and determine the structure of the function (e.g. its smoothness in frequency) which can be modelled with a GP. GPs are often described by parametrized priors in GPR, and the GP prior is selected such that it maximizes the Bayesian evidence, estimated by conditioning these priors to the observations (see Rasmussen & Williams 2005 for an extensive review). The parameters of the covariance functions (also known as ‘hyper-parameters’) can be estimated using standard optimization or Markov Chain Monte Carlo (MCMC) algorithms. The observed data  $\mathbf{d}$ , being a stacked set of gridded visibilities as a function of frequency, can be modelled as

$$\mathbf{d} = \mathbf{f}_{\text{fg}}(\nu) + \mathbf{f}_{21}(\nu) + \mathbf{n}, \quad (2)$$

where  $\mathbf{f}_{\text{fg}}(\nu)$  corresponds to the foreground component,  $\mathbf{f}_{21}(\nu)$  corresponds to the signal of interest, and  $\mathbf{n}$  is the noise. The 21-cm signal is expected to decorrelate over frequency scales  $> 1$  MHz, whereas foregrounds are expected to be smooth on 1 MHz scales and decorrelate over a much larger frequency range. The covariance function  $\mathbf{K} \equiv \kappa$  for this model can be written as a sum of foreground covariance function  $\mathbf{K}_{\text{fg}}$  and a 21-cm signal covariance function  $\mathbf{K}_{21}$ , i.e.

$$\mathbf{K} = \mathbf{K}_{\text{fg}} + \mathbf{K}_{21}. \quad (3)$$

$\mathbf{K}_{\text{fg}}$  can further be decomposed into  $\mathbf{K}_{\text{int}}$ , which corresponds to intrinsic foregrounds (large-scale correlation of  $\sim 10-100$  MHz) and  $\mathbf{K}_{\text{mix}}$ , which corresponds to instrumental mode mixing such as side-lobe noise (decorrelates within  $\sim 1 - 5$  MHz). The joint probability distribution for the observed data  $\mathbf{d}$  and function values  $\mathbf{f}_{\text{fg}}$  of the foreground model at the same frequency  $\nu$  is then given by

$$\begin{bmatrix} \mathbf{d} \\ \mathbf{f}_{\text{fg}} \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{\text{fg}} + \mathbf{K}_{21} + \sigma_n^2 \mathbf{I} & \mathbf{K}_{\text{fg}} \\ \mathbf{K}_{\text{fg}} & \mathbf{K}_{\text{fg}} \end{bmatrix} \right), \quad (4)$$

where  $\sigma_n^2$  is the noise variance,  $\mathbf{I}$  is the identity matrix, and  $\mathbf{K} \equiv \mathbf{K}(\nu, \nu)$ . After GPR, the foreground model can be retrieved as

$$E(\mathbf{f}_{\text{fg}}) = \mathbf{K}_{\text{fg}} [\mathbf{K} + \sigma_n^2 \mathbf{I}]^{-1} \mathbf{d}, \quad (5)$$

$$\text{cov}(\mathbf{f}_{\text{fg}}) = \mathbf{K}_{\text{fg}} - \mathbf{K}_{\text{fg}} [\mathbf{K} + \sigma_n^2 \mathbf{I}]^{-1} \mathbf{K}_{\text{fg}}, \quad (6)$$

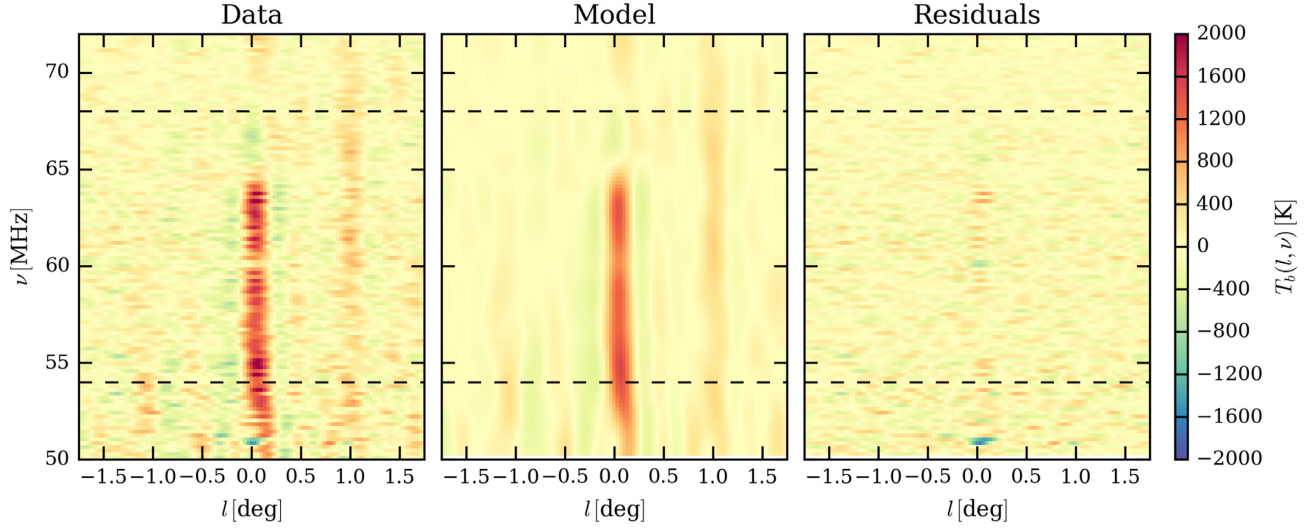
where  $E(\mathbf{f}_{\text{fg}})$  and  $\text{cov}(\mathbf{f}_{\text{fg}})$  are the expectation values and covariance of the foregrounds, respectively. The residuals  $\mathbf{d}_{\text{res}}$  after foreground model subtraction are

$$\mathbf{d}_{\text{res}} = \mathbf{d} - E(\mathbf{f}_{\text{fg}}). \quad (7)$$

Readers may refer to Mertens et al. (2018) for a detailed description of the GPR technique and its application as a novel method for foreground removal and 21-cm signal estimation.

### 5.2 Application of GPR to the LOFAR-LBA data

We use GPR to remove remaining foreground emission from the residual visibilities after DD calibration. We test various covariance functions as kernels to model different components of the residual visibilities in GPR. We use a Bayesian framework to compare different covariance functions and select those models that maximize



**Figure 7.** The 3C220 field Stokes  $I$  image cube slices (in brightness temperature units) across the centre of a spatial axis after different processing steps. Left-hand panel: a slice of the image cube after DD calibration (data). Middle panel: the GPR model of the smooth foregrounds (intrinsic + mode-mixing) corresponding to the data. Right-hand panel: the residuals after subtracting the GPR model from the data. The dashed black lines represent the frequency range (54–68 MHz) used for power spectrum estimation. The residuals after GPR are featureless except for a few outliers.

the marginal likelihood (or Bayesian evidence). We tested several GPR covariance kernels e.g. radial basis functions (RBF), rational quadratic (RQ) functions, and different classes of Matern covariance functions to model foreground components and finally selected the ones with the maximum Bayesian evidence.

To model the intrinsic foreground emission (unmodelled sources and diffuse emission) we choose an RBF covariance function as kernel. The RBF kernel is essentially a square exponential or Gaussian function defined as:

$$\kappa_{\text{RBF}}(\nu_p, \nu_q) = \exp\left(\frac{-|\nu_q - \nu_p|^2}{2l^2}\right) \quad (8)$$

where  $l$  is the characteristic coherence scale in frequency. We use 5–100 MHz as the prior for the range of coherence scales of the intrinsic foregrounds. To model the mode-mixing component of the foregrounds, we choose an RQ covariance function defined as:

$$\kappa_{\text{RQ}}(\nu_p, \nu_q) = \left(1 + \frac{|\nu_q - \nu_p|^2}{2\alpha l}\right)^{-\alpha}, \quad (9)$$

where  $l$  is the coherence scale and  $\alpha$  is the so-called power parameter. RQ functions can be understood as infinite sums of Gaussian covariance functions with characteristic coherence scales (Rasmussen & Williams 2005). We use 1–8 MHz as prior values for the coherence scales and  $\alpha = 0.1$  for the mode-mixing component. To account for the 21-cm signal, we use an exponential covariance function, which is a special case of a Matern class covariance function (Stein 1999) defined as:

$$\kappa_{\text{matern}}(\nu_p, \nu_q) = \frac{2^{1-n}}{\Gamma(n)} \left[\frac{\sqrt{2n}|\nu|}{l}\right]^n K_n\left(\frac{\sqrt{2n}|\nu|}{l}\right), \quad (10)$$

where  $|\nu| = |\nu_q - \nu_p|$  and  $K_n$  is the modified Bessel function of the second kind (not to be confused with the covariance functions defined in Section 5.1). The ‘hyper-parameter’  $l$  represents the characteristic coherence scale. Special classes of Matern covariance functions can be obtained by choosing various values for  $n$ , e.g. choosing  $n = 1/2$  gives an exponential kernel. We use a frequency coherence scale of 0.01–1.5 MHz for the 21-cm signal with an

initial value of 0.5 MHz. Using 21CMFAST simulations (Mesinger & Furlanetto 2007; Mesinger, Furlanetto & Cen 2011), Mertens et al. (2018) have shown that these coherence scales covers a wide range of possible 21-cm signal models.

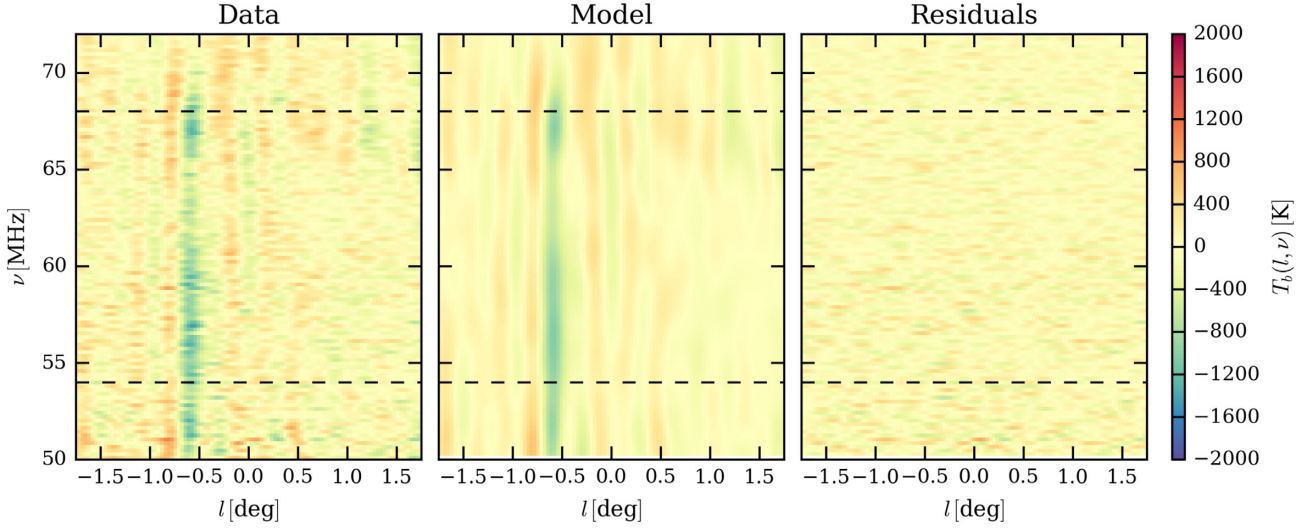
We use the residual image cubes obtained after DD calibration for foreground removal. We perform GPR foreground removal on the inner  $3.5^\circ \times 3.5^\circ$  region of the image cubes (which is slightly smaller than the primary beam FWHM  $\sim 4^\circ$ ) to limit sky curvature and primary beam effects. We selected the 50–72 MHz frequency range for GPR foreground removal, which is 8 MHz wider than the power spectrum estimation window, for better foreground fitting and removal. Fig. 7 shows slices through the Stokes  $I$  image cubes for the 3C220 field across the centre of one of the two spatial axes before GPR (left-hand panel), the GPR foreground fit (middle panel), and the residuals after subtracting the foreground model obtained with GPR from the data (right-hand panel). Similarly, Fig. 8 shows the slices of Stokes  $I$  image cubes for the NCP field. We observe that the Stokes  $I$  residuals after foreground removal with GPR for both the 3C220 and NCP fields are featureless (except for a few outliers) and do not appear to have spatial or spectral structure. In the following section, we use these residuals after GPR to create cylindrically and spherically averaged power spectra for the 3C220 and the NCP fields.

## 6 POWER SPECTRA RESULTS

In this section, we present the cylindrically and spherically averaged power spectra for the 3C220 and NCP fields. Cylindrically averaged power spectra (or 2D cosmological power spectra) are widely used in 21-cm experiments to assess various 21-cm signal contaminants such as foregrounds, side-lobe noise, and systematic biases. Cylindrically averaged power spectra ( $P(k_\perp, k_\parallel)$ ) are defined as (Parsons et al. 2012; Thyagarajan et al. 2015):

$$P(k_\perp, k_\parallel) = \frac{X^2 Y}{\Omega_A B} \langle |\tilde{\mathcal{V}}(\mathbf{u}, \eta)|^2 \rangle, \quad (11)$$





**Figure 8.** As Fig. 7 but for the NCP field. Similar to the 3C220 field, the residuals in the NCP field after GPR are featureless.

where  $\tilde{\mathcal{V}}(\mathbf{u}, \eta)$  is the FT of the visibilities in the frequency direction,  $\Omega_A$  is the integral of the square of the primary beam over solid angle across the sky, and  $B$  is the bandwidth of the visibility cube.  $X$  and  $Y$  are the conversion factors from angle and frequency to transverse comoving distance ( $D(z)$ ) and the comoving depth along the line of sight ( $\Delta D$ ), respectively. The wave numbers  $k_\perp$  and  $k_\parallel$  are related to baseline vector ( $\mathbf{u} = (u, v)$  in units of wavelength) and the Fourier dual to frequency ( $\eta$ ) as:

$$k_\perp = \frac{2\pi(|\mathbf{u}|)}{D(z)}, \quad k_\parallel = \frac{2\pi\nu_{21}H_0E(z)}{c(1+z)^2}\eta, \quad (12)$$

where  $\nu_{21}$  is the rest-frame frequency of the 21-cm spin flip transition of HI,  $z$  is the redshift corresponding to the observation frequency,  $c$  is the speed of light,  $H_0$  is the Hubble constant, and  $E(z) \equiv [\Omega_M(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda]^{1/2}$  is a function of the standard cosmological parameters. The spherically averaged dimensionless power spectrum can be estimated from  $P(k_\perp, k_\parallel)$  as:

$$\Delta^2(k) = \frac{k^3}{2\pi^2} P(k), \quad (13)$$

where  $k = \sqrt{k_\perp^2 + k_\parallel^2}$ . We determine  $P(k_\perp, k_\parallel)$  for both the 3C220 and NCP fields using the gridded visibility cubes of the  $3.5^\circ \times 3.5^\circ$  region of the image cubes with 14 MHz bandwidth (54–68 MHz), corresponding to the redshift range  $z = 19.8$ – $25.2$ . We weigh the  $uv$ -cells using an empirical weighting scheme where  $uv$ -cells in Stokes  $I$  and  $V$  with higher noise are downweighted. In this scheme, the weights are scaled by the inverse of the per-visibility variance  $\sigma^2[\mathcal{V}(u, v)]$  in each  $uv$ -cell estimated by computing robust variance statistics of Stokes  $V$  along the frequency direction as:

$$\sigma^2[\mathcal{V}(u, v)] = \{\hat{\sigma}_v[\sqrt{N_{\text{vis}}(u, v)} \times \mathcal{V}_V(u, v, \nu)]\}^2 \quad (14)$$

where  $\mathcal{V}_V(u, v)$  are the gridded Stokes  $V$  visibilities,  $N_{\text{vis}}(u, v, \nu)$  is the number of visibilities in a  $uv$ -cell, and  $\hat{\sigma}_v$  is a robust standard deviation estimator. The weights ( $W$ ) are then given by:

$$W(u, v, \nu) = N_{\text{vis}}(u, v, \nu) \frac{\langle \sigma^2[\mathcal{V}(u, v)] \rangle}{\sigma^2[\mathcal{V}(u, v)]}. \quad (15)$$

While the per-visibility variance should theoretically be identical in each  $uv$ -cell, it becomes different in the presence of systematics that

can affect baselines differently. This empirical weighting scheme allows one to reduce the impact of those systematics. We emphasize that only Stokes  $V$  is used in determining those weights.

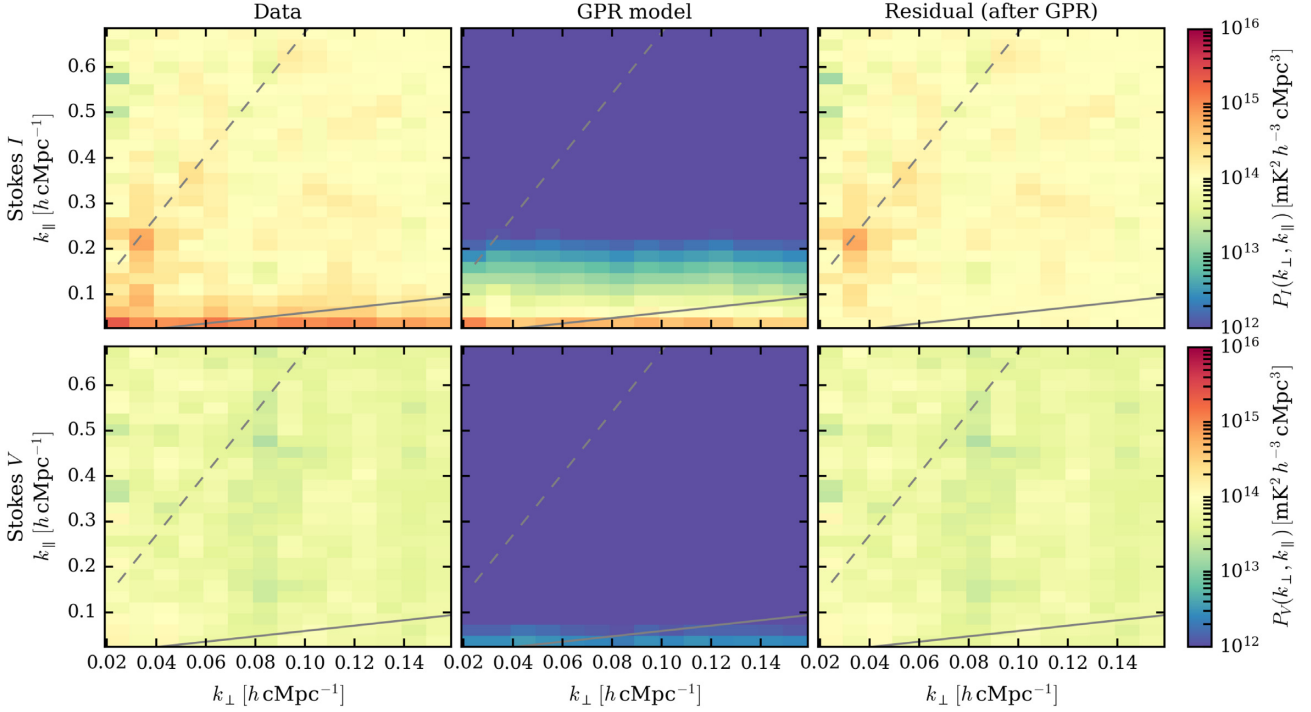
To FT the visibilities along the frequency direction, we use a Least-Squares Spectral Analysis method (full least-squares FT-matrix inversion, see e.g. Barning 1963; Lomb 1976; Stoica, Li & He 2009; Trott et al. 2016). We apply a ‘Hann’<sup>17</sup> window function to the data prior to the frequency transform to control the side-lobes along the  $\eta$  axis, however, this window function somewhat increases the noise. Although using a ‘Hann’ window introduces minor correlations between different  $k_\parallel$  modes, the residual spectra are similar to the ones produced using a top-hat window. Therefore, we currently ignore this effect in our analysis. The resulting  $\tilde{\mathcal{V}}(\mathbf{u}, \eta)$  cubes after frequency transform are scaled accordingly with the conversion factors  $X$  and  $Y$  calculated using  $\Lambda$  cold dark matter cosmological parameters that are consistent with the Planck 2016 results (Planck Collaboration et al. 2016), and then cylindrically and spherically averaged to obtain  $P(k_\perp, k_\parallel)$  and  $\Delta^2(k)$ , respectively.

### 6.1 The 3C220 field: cylindrical power spectra

In this section, we examine the cylindrical power spectra for the 3C220 field. The top row of Fig. 9 shows  $P(k_\perp, k_\parallel)$  for Stokes  $I$  before foreground removal, the GPR foreground model, and after GPR foreground removal. We observe that the lowest  $k_\parallel$  bins in Stokes  $I$  are dominated by smooth foreground emission (see top left panel) even after subtraction of the sky model during DD calibration. This foreground emission is modelled (shown in top middle panel) and effectively removed by the GPR foreground removal method (see top right panel). We also observe an excess power around the horizon line in Stokes  $I$  prior to GPR, which is not removed during GPR, suggesting that this excess power has much lower spectral smoothness or decorrelates quickly over time between gridded visibilities and cannot be modelled with a GP with current settings. This structure is reminiscent of the ‘pitchfork’ observed

<sup>17</sup>The ‘Hann’ window is defined as  $W(n) = 0.5 - 0.5 \cos[\frac{2\pi n}{(M-1)}]$ , where  $0 \leq n \leq M-1$  (see e.g. Blackman & Tukey 1958; Harris 1978).





**Figure 9.** The cylindrically averaged Stokes  $I$  and  $V$  power spectra for the 3C220 field. Top row (left to right):  $P_I(k_\perp, k_\parallel)$  before foreground removal, GPR foreground model, and after foreground removal with. Bottom row (left to right): same as top row, but for Stokes  $V$ . The solid grey lines correspond to a  $5^\circ$  FoV which is slightly larger than the primary beam FWHM at 60 MHz. The dashed grey lines correspond to the instrumental horizon.

in G18 and is possibly caused by the residuals after Cas A and Cyg A subtraction. An inaccurate source model, ionospheric effects, the time variation of the primary beam, or a combination of these effects might explain these residuals. We expect the modelling errors to be negligible as their corresponding models are derived from high spatial resolution images and also because Cas A and Cyg A appear as compact sources on shorter baselines. Ionospheric effects, however, become stronger at lower elevations due to projection effects and subtraction of Cas A and Cyg A at 30 s and 61 kHz calibration resolution might not be sufficient to correct for ionospheric effects, especially on the shorter baselines. Also, the primary beam changes with time as the 3C220 field is tracked. Therefore, a combination of ionospheric effects, beam errors and time variation of the primary beam is likely capable of producing such an effect.

The bottom row of Fig. 9 shows  $P(k_\perp, k_\parallel)$  for Stokes  $V$  before foreground removal, the GPR foreground model, and after GPR foreground removal. We observe that the Stokes  $V$  power spectrum is featureless before and after foreground removal, which suggests that any foreground emission/leakage in Stokes  $V$  is lower than the excess variance in the current data (see bottom middle panel). We also do not observe any visible signature of Cas A and Cyg A residuals. The vertical bands in Stokes  $I$  and  $V$  near  $k_\perp \approx 0.08$  and  $0.14$  are due to varying  $uv$ -density and drop out in the ratio  $\frac{P_I}{P_V}$ . The ratio after foreground removal, as shown in the right-hand panel of Fig. 10, is relatively flat compared to the one before foreground removal, except for the above-mentioned region near the horizon. The ratio has a median value of 2.07 which is higher than the median value ( $\sim 1.46$ ) observed in the ratio  $P_{\Delta I}/P_{\Delta V}$  for the 3C220 field (see Section 4.1). However, it is consistent with the excess at the sub-band level (see Section 4.2) caused by the use of a baseline cut in the DD calibration.

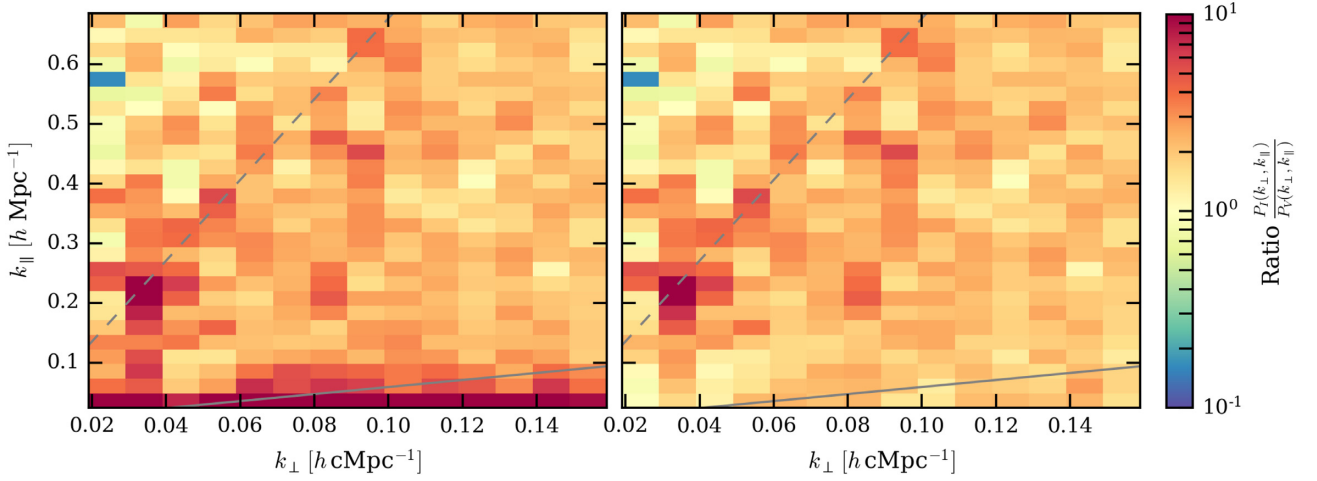
## 6.2 The NCP field: cylindrical power spectra

In this section, we assess the cylindrical power spectra for the NCP field. Fig. 11 shows  $P(k_\perp, k_\parallel)$  for Stokes  $I$  and  $V$ . The top left panel shows the power spectrum after DD calibration, where low  $k_\parallel$  modes are dominated by the power due to foreground emission. Similar to the 3C220 field, this power is effectively removed with GPR (see top right panel). We do not observe a ‘pitchfork’ in Stokes  $I$  or  $V$  (presumably) due to Cas A and Cyg A residuals opposed to the 3C220 field. This might be primarily because the NCP is stationary on the sky and therefore the beam does not change (only rotates) as the observation progresses. It is also likely that the Cas A and Cyg A are closer to the null for the NCP, causing the power on/around the structure to be below the noise. Similar to the 3C220 field, Stokes  $V$  power spectra for the NCP field appear featureless before and after foreground removal (see Fig. 11).

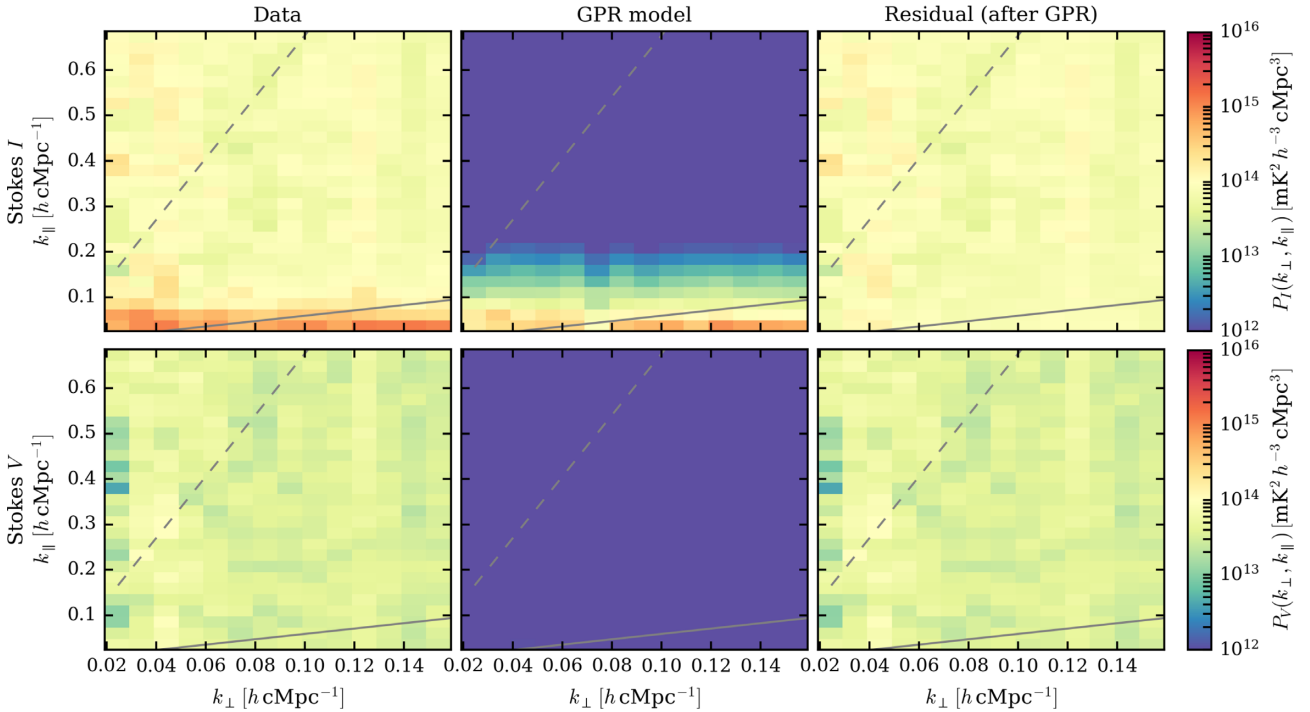
The behaviour of the ratio  $\frac{P_I}{P_V}$  (see Fig. 12) is also similar to that of the 3C220 field. The ratio becomes relatively flat after foreground removal except for a few outliers at the small  $k_\perp$  values. The ratio has a median value of 2.10, which is almost equivalent to the median we observed for the 3C220 field, but higher than the median of the ratio  $P_{\Delta I}/P_{\Delta V}$  for the NCP field (Section 4.1). This excess can also be attributed to the baseline cut in DD calibration as discussed in Section 4.2, which we know causes excess power.

## 6.3 Comparison with noise power spectra

We determine the cylindrically averaged noise power spectra  $P_I^N$  and  $P_V^N$  corresponding to the Stokes  $I$  and  $V$  difference cubes  $\Delta I$  and  $\Delta V$ , respectively, for the 3C220 field (e.g. see Section 4), by passing these cubes through the power spectrum estimation pipeline. Note that we do not perform foreground removal on these data cubes



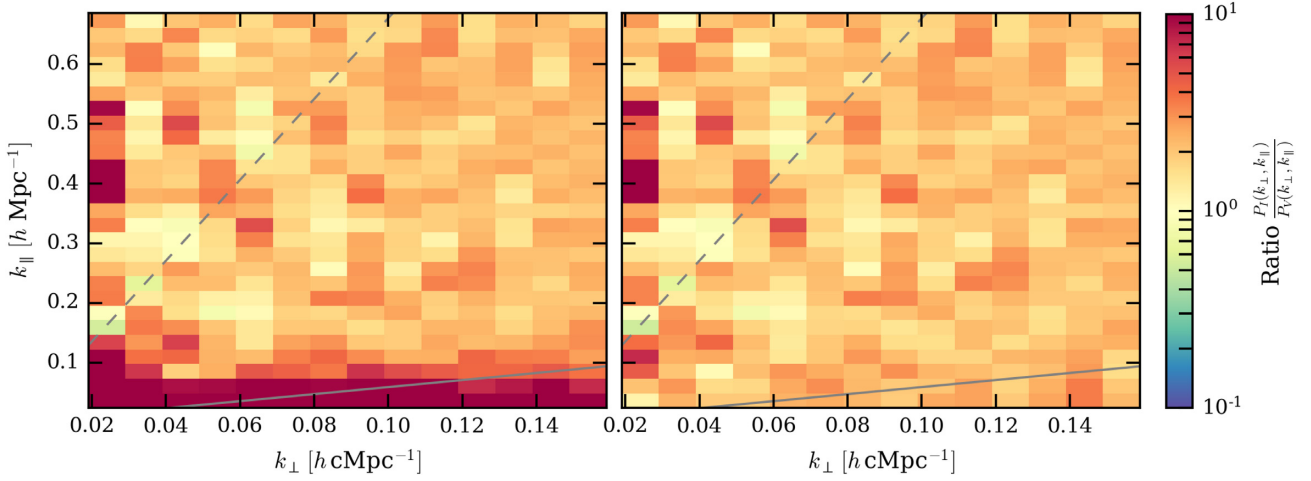
**Figure 10.** The ratio of the cylindrically averaged Stokes  $I$  and  $V$  power spectra for the 3C220 field. Left-hand panel:  $P_I/P_V$  before foreground removal with GPR. Right-hand panel:  $P_I/P_V$  after foreground removal with GPR.



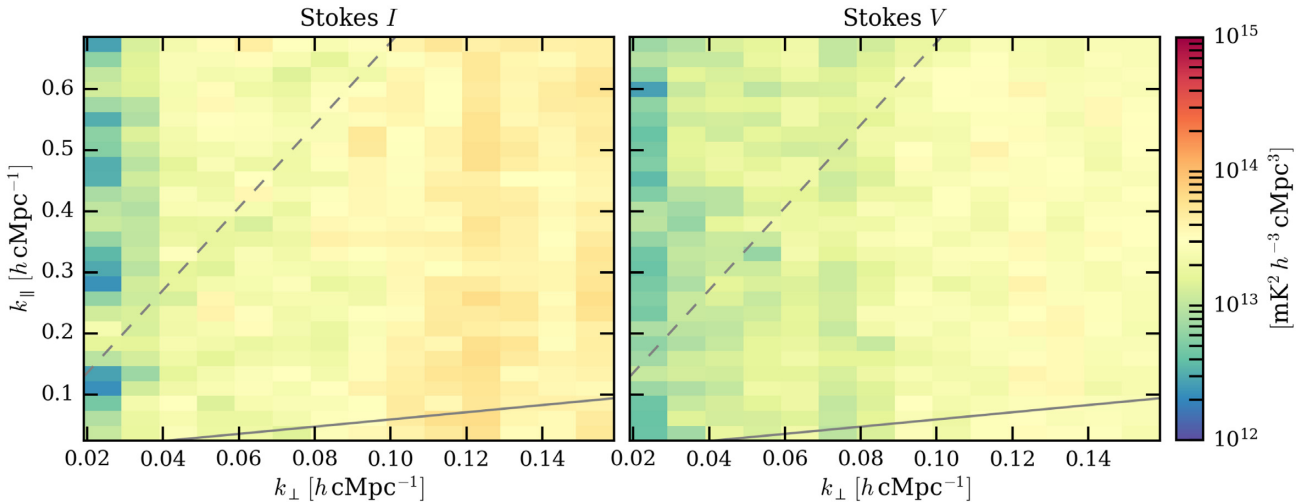
**Figure 11.** The cylindrically averaged Stokes  $I$  and  $V$  power spectra for the NCP field. Top row (left to right):  $P_I(k_\perp, k_\parallel)$  before foreground removal, GPR foreground model, and after foreground removal with. Bottom row (left to right): same as top row, but for Stokes  $V$ .

because we expect the sky component to dropout on time-scales of 12 s. Fig. 13 shows  $P_I^N$  (left-hand panel) and  $P_V^N$  (right-hand panel). We observe that power in both Stokes  $I$  and  $V$  is lower for small  $k_\perp$  values and higher for larger  $k_\perp$  because the  $uv$ -density of LOFAR-LBA decreases with increasing baseline length and drops out in the ratio  $\frac{P_I^N}{P_V^N}$  shown in Fig. 14. From comparison of  $P_V(k_\perp, k_\parallel)$  for the 3C220 and NCP fields with  $P_V^N$ , we notice that  $P_V(k_\perp, k_\parallel)$  deviates from  $P_V^N$  for lower  $k_\perp$  ( $<0.1$ ) values. This deviation in  $P_V(k_\perp, k_\parallel)$  compared to  $P_V^N$  can be attributed to the baseline cut used in the DD calibration, which increases the noise on the baselines excluded from the calibration.

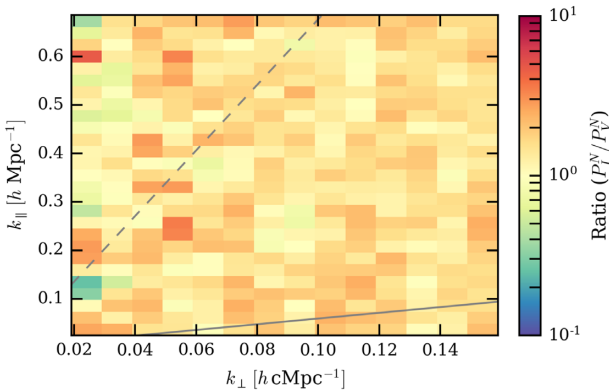
Moreover,  $\frac{P_I^N}{P_V^N}$  has a median value of 1.51, which is consistent with the median value of 1.46 for the ratio  $P_{\Delta I}/P_{\Delta V}$ . The NCP field, however, has a slightly lower median value of 1.3. We observe that this excess power in Stokes  $I$  for both the 3C220 and the NCP fields at 12 s level does not depend on the calibration, and is present at the same level throughout the analysis even after DD calibration, foreground removal, and also in the autocorrelations (results not shown here). This excess is different from the calibration cut induced noise and might have a physical origin. To account for this physical excess noise in the estimation of the spherically averaged power spectrum, we multiply the residual Stokes  $V$  gridded



**Figure 12.** The ratio of the cylindrically averaged Stokes  $I$  and  $V$  power spectra for the NCP field. Left-hand panel:  $P_I/P_V$  before foreground removal with GPR. Right-hand panel:  $P_I/P_V$  after foreground removal with GPR.



**Figure 13.** The cylindrically averaged Stokes  $I$  and  $V$  noise power spectra  $P_I^N$  (left-hand panel) and  $P_V^N$  (right-hand panel) for the 3C220 field determined from the difference cubes  $\Delta_I \tilde{I}$  and  $\Delta_I \tilde{V}$ , respectively.

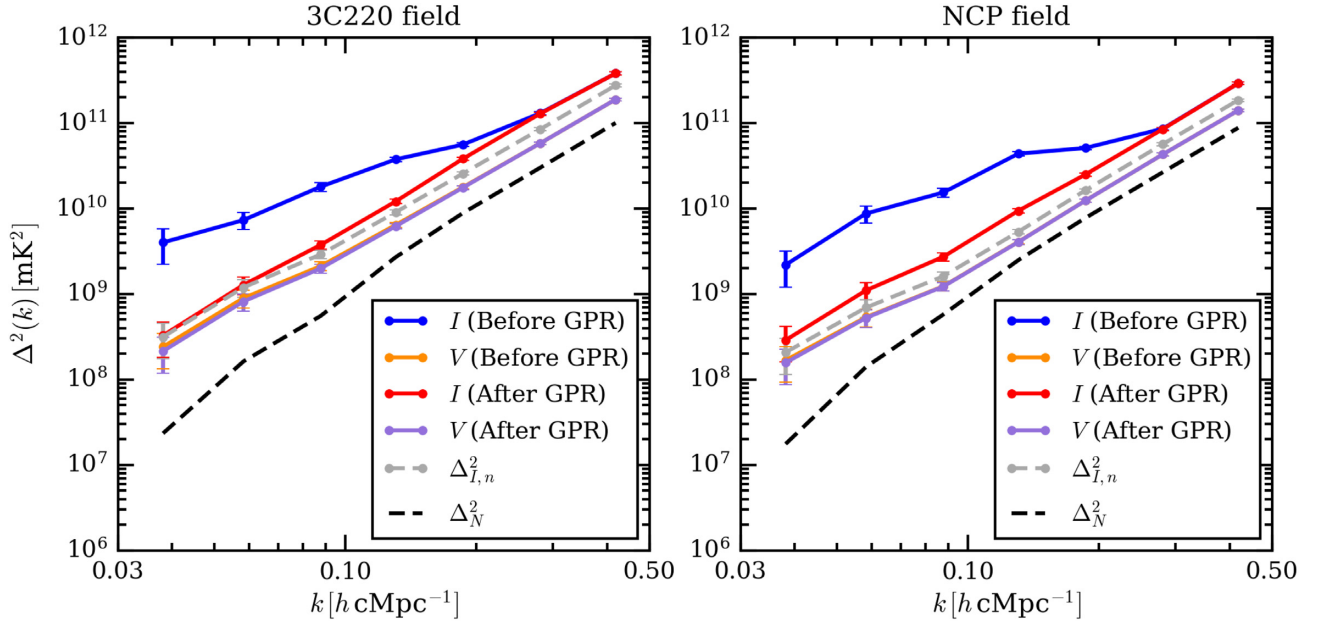


**Figure 14.** The ratio  $\frac{P_I^N}{P_V^N}$  for the 3C220 field. We observe that the ratio has a median value of 1.51.

visibilities after DD calibration (since Stokes  $V$  is an independent estimator of the thermal noise of the instrument) with the square root of the median of the ratio  $P_{\Delta_I I} / P_{\Delta_I V}$  (calculated in Section 4.1) to obtain an estimate of the noise in Stokes  $I$ . We use the median instead of the mean because the median is a more robust representative of the central tendency of the distribution, whereas the mean is sensitive to outliers and becomes biased in the presence of strong outliers. This excess noise bias corrected Stokes  $V$  is used as an estimator for the noise in the data in the foreground removal and spherically averaged power spectrum estimation steps.

#### 6.4 Spherically averaged power spectra

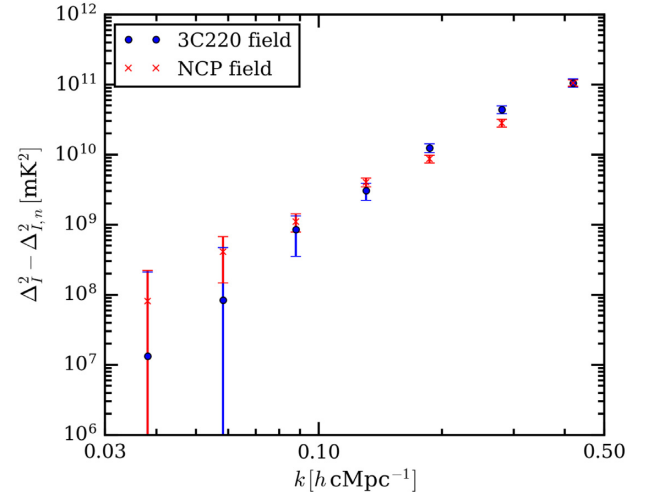
We finally determine the Stokes  $I$  and  $V$  spherically averaged power spectra ( $\Delta^2(k)$ ) for both the 3C220 and NCP fields before and after foreground removal for the redshift range  $z = 19.8$ – $25.2$ . Fig. 15 shows Stokes  $I$  power spectra  $\Delta_I^2$  and Stokes  $V$  power spectra  $\Delta_V^2$  for both the 3C220 (left-hand panel) and NCP fields (right-hand panel)



**Figure 15.** The spherically averaged Stokes  $I$ ,  $V$ , and excess noise bias corrected Stokes  $V$  power spectra. Left-hand panel:  $\Delta_I^2$  and  $\Delta_V^2$  for the 3C220 field before (blue and orange curves, respectively) and after (red and purple curves, respectively) foreground removal. Right-hand panel:  $\Delta_I^2$  and  $\Delta_V^2$  for the NCP field using the same colour scheme as in the left-hand panel. The dashed grey and dashed black curves represent noise bias corrected Stokes  $V$  power spectrum  $\Delta_{I,n}^2$  and noise power spectrum estimate  $\Delta_N^2$ , respectively, for the corresponding fields. The error bars represent the  $2\sigma$  errors on the power spectra.

before and after foreground removal. We use the physical excess noise bias corrected (using the median values from Section 4.1) Stokes  $V$  visibilities as an estimator of the noise component in the Stokes  $I$  power spectrum ( $\Delta_{I,n}^2$ ), in order to account for the physical excess noise in Stokes  $I$  compared to Stokes  $V$ . The dashed grey curves in Fig. 15 represent the excess bias corrected Stokes  $V$  power spectra  $\Delta_{I,n}^2$ . For both fields, we observe that the power on smaller  $k$  modes is dominated by large-scale foreground emission which comprises diffuse emission, unmodelled sources, and sources below the confusion noise prior to foreground removal. A recent analysis of the wide-field AARTFAAC-12 HBA data (at 122 MHz) presented in Gehlot (2019) shows strong diffuse emission around the NCP on degree scales ( $u < 200$ ) which is well beyond the thermal noise. This emission becomes stronger at lower frequencies (50–70 MHz) suggesting that the smallest  $k$  modes are fully dominated by diffuse emission. Residual Stokes  $I$  power on the smallest  $k$  modes after foreground removal is an order of magnitude lower than the former. However, GPR does not remove any power from Stokes  $V$ , which means that any structure in Stokes  $V$  is spatially and spectrally incoherent and behaves as uncorrelated noise. For the 3C220 field, Stokes  $I$  residuals approach  $\Delta_{I,n}^2$  at smaller  $k$  modes, however these are still higher than  $\Delta_{I,n}^2$  by  $\sim 30$  per cent on large  $k$  modes. On the other hand, Stokes  $I$  residuals for the NCP field are higher than  $\Delta_{I,n}^2$  by  $\sim 50$  per cent on most  $k$  modes except for the lowest one. This remaining excess power, after correcting for the physical excess noise bias, is likely due to the baseline cut used during the DD calibration.

Assuming that the physical noise properties of Stokes  $I$  and  $V$  are statistically identical, we use the post GPR excess noise bias corrected Stokes  $V$  power spectrum ( $\Delta_{I,n}^2$ ) to remove the noise component in the residual Stokes  $I$  power spectrum. The noise bias corrected power spectrum  $\Delta_I^2 - \Delta_{I,n}^2$  for the 3C220 (blue circles) and the NCP field (red crosses) are shown in Fig. 16. The dashed curves show thermal noise power spectrum estimate  $\Delta_N^2$  estimated



**Figure 16.** Noise bias corrected spherically averaged Stokes  $I$  power spectra ( $\Delta_I^2 - \Delta_{I,n}^2$ ) for the 3C220 and NCP fields. Blue circles represent the 3C220 field and red crosses represent the NCP field. The error bars correspond to the  $2\sigma$  errors on the power spectra.

from  $\Delta_I \tilde{V}$  for the 3C220 (‘skyblue’ coloured) and the NCP field (‘coral’ coloured). We observe that  $\Delta_I^2 - \Delta_{I,n}^2$  for both fields are consistent with each other within the  $2\sigma$  uncertainty for modes  $k \lesssim 0.2 h cMpc^{-1}$  and deviate slightly on  $k \gtrsim 0.2 h cMpc^{-1}$ . This is possibly due to different morphologies of the two fields on small angular scales. The  $\Delta_I^2 - \Delta_{I,n}^2$  for both fields, within  $2\sigma$  uncertainty, agree with their respective noise estimate  $\Delta_N^2$  (determined from  $\Delta_I \tilde{V}$ ) which is a more accurate estimator of the thermal noise of the system. The  $\Delta_N^2$  for both fields show power-law-like behaviour and agree with each other on all  $k$  modes. We find a  $2\sigma$  upper limit



of  $\Delta_{21}^2 < (14561 \text{ mK})^2$  at  $k \sim 0.038 h \text{ cMpc}^{-1}$  for the 3C220 field and  $\Delta_{21}^2 < (14886 \text{ mK})^2$  at  $k \sim 0.038 h \text{ cMpc}^{-1}$  for the NCP field in the redshift range  $z = 19.8\text{--}25.2$ . Both upper limits are consistent with each other within 2 per cent. The upper limits  $\Delta_1^2 - \Delta_{1,n}^2$  for the two fields are still dominated by systematics on most  $k$  modes. A deeper understanding of systematics (and their mitigation) and a more accurate estimate of the noise bias is required to remove this additional bias. From our current analysis, we observed that it is harder to model the exact noise bias, which is crucial to obtain more robust upper limit. We are currently developing improved estimators of the incoherent noise power spectrum which might be thermal noise and also include incoherent random errors e.g. due to the ionosphere, calibration, etc., for noise bias subtraction. We are also exploring other methods (e.g. cross-correlating independent data sets) to estimate 21-cm power spectrum which circumvents several issues with noise bias subtraction and plan to incorporate them in future analyses.

## 7 SUMMARY AND OUTLOOK

In this work, we have explored the possibility of statistical measurement of the redshifted 21-cm signal of HI from the CD using the LOFAR-LBA system. We have presented the first upper limits on the power spectrum of the 21-cm signal in the high-redshift range of  $z = 19.8\text{--}25.2$  using LOFAR-LBA data with dual-pointing setup pointed at the NCP and the radio galaxy 3C220.3 simultaneously. Our main conclusions are:

(i) For the 3C220 field, after 14 h of integration, a  $2\sigma$  upper limit of  $\Delta_{21}^2 < (14561 \text{ mK})^2$  at  $k = 0.038 h \text{ cMpc}^{-1}$  is reached on the power spectrum of 21-cm brightness temperature fluctuations. Similarly, for the NCP field, we reach a  $2\sigma$  upper limit of  $\Delta_{21}^2 < (14886 \text{ mK})^2$  at  $k = 0.038 h \text{ cMpc}^{-1}$  in the redshift range  $z = 19.8\text{--}25.2$ . Both upper limits are consistent with each other within 2 per cent level. Upper limits for both the 3C220 and the NCP fields are still dominated by the systematics.

(ii) We demonstrate the application of a multiple pointing method to calibrate LOFAR-LBA dual-pointing observations.

(iii) We observe an excess of noise in the ratio of Stokes  $I$  and  $V$  noise spectra over short time-scales (12 s) in baseline-frequency space, derived from the Stokes  $I$  and  $V$  difference image cubes created from even and odd visibility samplings at 12-s level. This excess is independent of frequency and baseline length and is also not affected by calibration. This excess noise is different from that introduced during calibration and already exists before DI and after DD calibration and does not change during those steps. The excess is different for the two fields and seems to have no clear origin. We suspect it to be caused by (diffractive) ionospheric scintillation noise, but we leave this analysis for the future.

(iv) We show that introducing frequency smoothness of instrumental gains as a constraint in both DI and DD calibration of LOFAR-LBA data greatly reduces the calibration induced excess variance on the sub-band level in Stokes  $I$  compared to Stokes  $V$  in contrast to the unconstrained case presented in G18, where we found an excess by a factor  $\sim 10$ . However, an excess of  $\sim 2\text{--}3$  still remains, which can be explained by the exclusion of short baselines during DD calibration as shown for LOFAR-HBA data calibration as well in Patil et al. (2016) and Mouri Sardarabadi & Koopmans (2018).

(v) After foreground removal using GPR, the Stokes  $I$  power spectrum is  $\sim 2$  times that of Stokes  $V$  for both fields and is featureless on most scales. However, we observe a ‘pitchfork’ like

structure in the 3C220 field at low  $k_\perp$  near the horizon line. We expect this structure to be caused by Cas A and Cyg A residuals as seen by G18.

## 7.1 Outlook

Detection of the redshifted 21-cm HI signal from CD and the EoR promises to be an excellent probe to study the early phases of the evolution of the Universe and has the potential to unveil exotic astrophysical phenomena. With the analysis shown in this work, a CD experiment with LOFAR-LBA will require  $>10^4$  h of integration (power spectrum sensitivity of  $\sim (100 \text{ mK})^2$  in CD redshift range) in order to constrain the optimistic CD X-ray heating and baryon–dark matter scattering models (see e.g. Fialkov, Barkana & Cohen 2018; Cohen, Fialkov & Barkana 2018). We plan to improve the analysis in the future by improving the enforcement of spectral smoothness in calibration, better modelling of the instrument (improving beam models), and by using the new Image Domain Gridded combined with WSCLEAN (see e.g. Veenboer, Pletschow & Romein 2017; van der Tol et al. 2018) to subtract off-axis sources. The upcoming LOFAR 2.0 upgrade will also increase the sensitivity of the LOFAR-LBA system. The combination of all these improvements will allow us to improve the CD 21-cm power spectrum sensitivity significantly.

Moreover, recently a deep absorption feature ( $\sim 0.5$  K deep) centred at  $\sim 78$  MHz ( $z \sim 17$ ) in the averaged sky spectrum was presented by Bowman et al. (2018), possibly being the sought-after 21-cm signal absorption feature seen against the cosmic microwave background during the CD era. The suggested absorption feature is considerably ( $\sim 2.5$  times) stronger and wider than predicted by standard astrophysical models (Barkana 2018). If confirmed, such a strong signal will lead to a large increase in the 21-cm brightness temperature fluctuations in the redshift range  $z = 17\text{--}19$  corresponding to the deepest part of the absorption profile (Barkana 2018; Fialkov et al. 2018), making it possible to detect its signal in a much shorter integration time compared to what was previously expected. Motivated by this, we have commenced a large scale program called the AARTFAAC Cosmic Explorer which uses the Amsterdam-ASTRON Radio Transients Facility And Analysis Center (AARTFAAC) correlator based on LOFAR, to obtain wide-field data for statistical detection of the 21-cm brightness temperature fluctuations within the redshift range of the absorption feature. The techniques discussed in this paper and lessons learned here will be useful in understanding and mitigating the challenges in AARTFAAC data processing and analysis, as well as in the NenuFAR, the HERA, and the upcoming SKA-low, which can also observe the similar redshift range. The SKA-low will also support multibeam observations, and thus also can take advantage of the dual-beam calibration strategy we have demonstrated.

## ACKNOWLEDGEMENTS

BKG and LVEK acknowledge the financial support from a NOVA cross-network grant. FGM acknowledges support from a SKA-NL (Square Kilometer Array - Netherlands) roadmap grant from the Dutch Ministry of OCW (Ministerie van Onderwijs, Cultuur en Wetenschappen). LOFAR, the Low Frequency Array designed and constructed by ASTRON, has facilities in several countries, that are owned by various parties (each with their own funding sources), and that are collectively operated by the International LOFAR Telescope (ILT) foundation under a joint scientific policy.

## REFERENCES

- Ali Z. S. et al., 2015, *ApJ*, 809, 61
- Ali Z. S. et al., 2018, *ApJ*, 863, 201
- Barkana R., 2018, *Nature*, 555, 71
- Barning F. J. M., 1963, *Bull. Astron. Inst. Netherlands*, 17, 22
- Barry N., Hazelton B., Sullivan I., Morales M. F., Pober J. C., 2016, *MNRAS*, 461, 3135
- Beardsley A. P. et al., 2016, *ApJ*, 833, 102
- Bernardi G. et al., 2016, *MNRAS*, 461, 2847
- Blackman R. B., Tukey J. W., 1958, *Bell Syst. Tech. J.*, 37, 185
- Bowman J. D., Rogers A. E. E., Monsalve R. A., Mozdzen T. J., Mahesh N., 2018, *Nature*, 555, 67
- Bowman J. D. et al., 2013, *PASA*, 30, e031
- Boyd S., Parikh N., Chu E., Peleato B., Eckstein J., 2011, *Found. Trends Mach. Learn.*, 3, 1
- Cohen A., Fialkov A., Barkana R., 2018, *MNRAS*, 478, 2193
- Cohen A. S., Lane W. M., Cotton W. D., Kassim N. E., Lazio T. J. W., Perley R. A., Condon J. J., Erickson W. C., 2007, *AJ*, 134, 1245
- DeBoer D. R. et al., 2017, *PASP*, 129, 045001
- de Gasperin F., et al., 2019, *A&A*, 622, A5
- de Gasperin F., Mevius M., Rafferty D. A., Intema H. T., Fallows R. A., 2018, *A&A*, 615, A179
- Ewall-Wice A., Dillon J. S., Liu A., Hewitt J., 2017, *MNRAS*, 470, 1849
- Ewall-Wice A. et al., 2016, *MNRAS*, 460, 4320
- Fialkov A., Barkana R., 2014, *MNRAS*, 445, 213
- Fialkov A., Barkana R., Cohen A., 2018, *Phys. Rev. Lett.*, 121, 011101
- Furlanetto S. R., Oh S. P., Briggs F. H., 2006, *Phys. Rep.*, 433, 181
- Gehlot B., 2019, PhD thesis. University of Groningen
- Gehlot B. K. et al., 2018, *MNRAS*, 478, 1484
- Glover S. C. O., Brand P. W. J. L., 2003, *MNRAS*, 340, 210
- Hales S. E. G., Waldram E. M., Rees N., Warner P. J., 1995, *MNRAS*, 274, 447
- Harris F. J., 1978, *Proc. IEEE*, 66, 51
- Jacobs D. C. et al., 2016, *ApJ*, 825, 114
- Kazemi S., Yatawatta S., Zaroubi S., 2013, *MNRAS*, 430, 1457
- Koopmans L. et al., 2015, in Burke Tyler L. et al., eds. *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*. Proceedings of Science, Italy, p. 1
- Laing R. A., Peacock J. A., 1980, *MNRAS*, 190, 903
- Lomb N. R., 1976, *Ap&SS*, 39, 447
- Madau P., Meiksin A., Rees M. J., 1997, *ApJ*, 475, 429
- Mellema G. et al., 2013, *Exper. Astron.*, 36, 235
- Mertens F. G., Ghosh A., Koopmans L. V. E., 2018, *MNRAS*, 478, 3640
- Mesinger A., Furlanetto S., 2007, *ApJ*, 669, 663
- Mesinger A., Furlanetto S., Cen R., 2011, *MNRAS*, 411, 955
- Mouri Sardarabadi A., Koopmans L. V. E., 2018, *MNRAS*, 483, 5480
- Offringa A. R., de Bruyn A. G., Biehl M., Zaroubi S., Bernardi G., Pandey V. N., 2010, *MNRAS*, 405, 155
- Offringa A. R., van de Gronde J. J., Roerdink J. B. T. M., 2012, *A&A*, 539, A95
- Offringa A. R. et al., 2014, *MNRAS*, 444, 606
- Paciga G. et al., 2011, *MNRAS*, 413, 1174
- Paciga G. et al., 2013, *MNRAS*, 433, 639
- Parsons A. R., Pober J. C., Aguirre J. E., Carilli C. L., Jacobs D. C., Moore D. F., 2012, *ApJ*, 756, 165
- Parsons A. R. et al., 2010, *AJ*, 139, 1468
- Patil A. H. et al., 2016, *MNRAS*, 463, 4317
- Patil A. H. et al., 2017, *ApJ*, 838, 65
- Philip L. et al., 2018, *J. Astron. Instrum.*, 8, 1950004
- Planck Collaboration XLVII et al., 2016, *A&A*, 596, A108
- Pritchard J. R., Furlanetto S. R., 2007, *MNRAS*, 376, 1680
- Pritchard J. R., Loeb A., 2012, *Rep. Prog. Phys.*, 75, 086901
- Rasmussen C. E., Williams C. K. I., 2005, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, Cambridge, MA
- Ross H. E., Dixon K. L., Iliev I. T., Mellema G., 2017, *MNRAS*, 468, 3785
- Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, *A&A*, 345, A380
- Singh S. et al., 2017, *ApJ*, 845, L12
- Stein M., 1999, *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Series in Statistics. Springer, New York
- Stoica P., Li J., He H., 2009, *IEEE Trans. Signal Process.*, 57, 843
- Thyagarajan N. et al., 2015, *ApJ*, 804, 14
- Tingay S. J. et al., 2013, *PASA*, 30, e007
- Trott C. M. et al., 2016, *ApJ*, 818, 139
- van der Tol S., Veenboer B., Offringa A. R., 2018, *A&A*, 616, A27
- van Haarlem M. P. et al., 2013, *A&A*, 556, A2
- Veenboer B., Petschow M., Romein J. W., 2017, in Parashar M., ed., 2017 IEEE International Parallel and Distributed Processing Symposium (IPDPS). Conference Publishing Services (CPS) of the IEEE Computer Society, USA, p. 545
- Voytek T. C., Natarajan A., Jáuregui García J. M., Peterson J. B., López-Cruz O., 2014, *ApJ*, 782, L9
- Yatawatta S., 2011, in Banerjee P., ed., 2011 XXXth URSI General Assembly and Scientific Symposium. IEEE, Turkey, p. 1
- Yatawatta S., 2015, *MNRAS*, 449, 4506
- Yatawatta S., 2016, in Hanzo L., ed., 2016 24th European Signal Processing Conference (EUSIPCO). IEEE, Hungary, p. 265
- Yatawatta S., 2018, in Liu W., ed., 2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM). IEEE, UK, p. 485
- Yatawatta S., Dibley F., Spreuw H., 2017, in de Almeida A. L. F., ed., 2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). IEEE, Netherlands Antilles, p. 1
- Zarka P., Girard J. N., Tagger M., Denis L., 2012, in Boissier S., de Laverny P., Nardetto N., Samadi R., Valls-Gabaud D., Wozniak H., eds, SF2A-2012: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics. Société Française d'Astronomie et d'Astrophysique, France, p. 687
- Zaroubi S., 2013, in Wiklund T., Mobasher B., Bromm V., eds, *Astrophysics and Space Science Library* Vol. 396, The First Galaxies. Springer-Verlag, Berlin, p. 45

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.